BMJ Open

Predictors of Incident SARS-CoV-2 Infections in Participants of the Covid-19 Citizen Science Study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-052025
Article Type:	Original research
Date Submitted by the Author:	20-Apr-2021

Complete List of Authors:	Lin, Anthony; University of California San Francisco, Vittinghoff, Eric; University of California San Francisco, Epidemiology and Biostatistics Olgin, J; University of California San Francisco Peyser, Noah; University of California San Francisco Aung, Sidney; University of California San Francisco Joyce, Sean; University of California San Francisco Yang, Vivian; University of California San Francisco Hwang, Janet; University of California San Francisco Avram, Robert; University of California San Francisco Nah, Gregory; University of California San Francisco Tison, Geoffrey ; University of California San Francisco, Beatty, Alexis; University of California San Francisco, Runge, Ryan; University of California San Francisco Wen, David; University of California San Francisco Butcher, Xochitl; University of California San Francisco Horner, Cathy; University of California San Francisco Eitel, Helena; University of California San Francisco Pletcher, Mark; University of California San Francisco Marcus, GM; University of California San Francisco

Keywords:	COVID-19, PUBLIC HEALTH, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, Public health < INFECTIOUS DISEASES

SCHOLARONE™ Manuscripts

Predictors of Incident SARS-CoV-2 Infections in Participants of the Covid-19 Citizen Science Study

Anthony L Lin, MD,[1] Eric Vittinghoff, PhD,[2] Jeffrey E Olgin, MD,[3] Noah D Peyser, PhD,[3] Sidney Aung,[4] Sean Joyce,[3] Vivian Yang,[3] Janet Hwang,[3] Robert Avram, MD,[3] Gregory Nah, MA,[3] Geoffrey H Tison, MD, MPH,[3] Alexis L Beatty, MD, MAS,[3] Ryan Runge,[3] David Wen,[3] Xochitl Butcher,[3] Cathy Horner,[3] Helena Eitel,[3] Mark J Pletcher, MD, MPH,[2] Gregory M Marcus, MD, MAS[3]

Author Affiliations

[1]Department of Medicine, University of California, San Francisco

[2]Department of Epidemiology and Biostatistics, University of California, San Francisco

[3]Division of Cardiology, Department of Medicine, University of California, San Francisco

[4]School of Medicine, University of California, San Francisco

Corresponding Author

Gregory M Marcus, MD, MAS

505 Parnassus Ave, M1180B

San Francisco, CA 94143

Phone: 415-476-5706

Fax: 415-476-6260

Email: greg.marcus@ucsf.edu

Word Count: 2,900

ABSTRACT

Objective: Until effective treatments and vaccines are made readily and widely available, preventative behavioral health measures will be central to the SARS-CoV-2 public health response. While current recommendations are grounded in general infectious disease prevention practices, it is still not entirely understood which particular behaviors or exposures meaningfully affect one's own risk of incident SARS-CoV-2 infection. Our objective is to identify individual-level factors associated with one's personal risk for contracting SARS-CoV-2.

Design: Prospective cohort study of adult participants from March 26, 2020 to October 8, 2020.

Setting: The Covid-19 Citizen Science Study, an international, community and mobile-based study collecting daily, weekly, and monthly surveys in a prospective and time-updated manner.

Participants: All adult participants over the age of 18 years were eligible for enrollment.

Primary Outcome Measure: The primary outcome was incident SARS-CoV-2 infection confirmed via polymerase chain reaction or antigen testing.

Results: 28,575 unique participants contributed 2,479,149 participant-days of data across 99 different countries. Of these participants without a history of SARS-CoV-2 infection at time of enrollment, 112 developed an incident infection. Pooled logistic regression models showed that increased age was associated with lower risk (OR 0.98 per year, 95% CI 0.97-1.00, p=0.019), whereas increased number of non-household contacts (OR 1.10 per 10 contacts, 95% CI 1.01-1.20, p=0.024), attending events of at least 10 people (OR 1.26 per 10 events, 95% CI 1.07-1.50, p=0.007), and restaurant visits (OR 1.95 per 10 visits, 95% CI 1.42-2.68, p<0.001) were associated with significantly higher risk of incident SARS-CoV-2 infection.

Conclusions: Our study identified 3 modifiable health behaviors, including the number of non-household contacts, attending large gatherings, and restaurant visits that may meaningfully influence individual-level risk of contracting SARS-CoV-2.

Strengths and limitations of this study

- This large international cohort study with 2.4 million participant-days of data from participants in 99 different countries provides unprecedented geographical diversity for a study analyzing individual-level factors associated with SARS-CoV-2 risk.
- All participants included in this study were free of SARS-CoV-2 infection early in the pandemic, allowing for the real-time ascertainment of significant individual-level behaviors and exposures related to higher risk of incident infection.
- Using polymerase chain reaction or antigen testing as the gold standard for SARS-CoV-2 infections relied on a participant's development of symptoms, index of suspicion, and access to testing facilities, but ensured our study identified risk factors associated with true infection and increased specificity over traditional methods of symptom reporting.

INTRODUCTION

The novel coronavirus (SARS-CoV-2) global pandemic has created a major public health crisis for nearly every country and community in the world. Responses to mitigate transmission have varied by government, but have generally been grounded in known respiratory virus disease prevention practices. Current strategies have included a combination of social distancing, limitations to travel and public gatherings, increased handwashing practices, and the use of face masks. While these interventions are believed to reduce human-to-human transmission, efforts to study these interventions have been limited as they rely on individual-level behaviors that are dynamic with policy changes and can be difficult to capture at scale. Furthermore, the politicization of social distancing recommendations[1–3] makes it difficult to fully understand levels of compliance at the individual-level, and calls for a larger evidence base for recommendations like hand washing, face mask wearing, and limiting human contact, large social gatherings, and visits to restaurants. Identifying predictors of infection requires a longitudinal cohort study. The information gleaned from the longitudinal characterization of SARS-CoV-2 infection risk factors may be crucial to understanding which strategies are most effective and can further inform public policy. Moreover, such data may help elucidate the individual behaviors directly under one's control to influence one's personal risk of contracting SARS-CoV-2.

While previous prospective studies have focused primarily on symptom detection and the constellation of symptoms associated with SARS-CoV-2 infection,[4–7] mobile technology provides an opportunity to study the effects of various exposures and behaviors that can be ascertained prospectively, repeatedly, and in nearly real-time. The majority of previous research regarding SARS-CoV-2 has focused on hospitalized individuals, primarily those who already have the disease, and predictors of disease severity as opposed to those pertinent to developing infection. This is not surprising as accumulating sufficient numbers to characterize non-infected individuals at baseline and then follow them over time is generally time-consuming and would require enrollment of particularly large numbers to derive useful results. While systematic reviews and meta-analyses of previous studies have investigated the efficacy of behavioral interventions,[8,9] we are not aware of a longitudinal cohort study in which risk factors have been characterized in detail prior to infection and exposures and behaviors tracked as individuals contracted (or did not contract) SARS-CoV-2 in the community.

Given the widespread use of smartphones and associated mobile apps, the technology is now available to regularly query large populations to assess patterns in SARS-CoV-2 infection rates based on individual-level exposures and behaviors. We have previously demonstrated the utility of this technology in characterizing ambulatory cardiovascular risk factors.[10–14] In this study, we sought to use prospectively-collected information from the Covid-19 Citizen Science Study to identify individual characteristics, exposures, or behaviors associated with an increased risk of contracting SARS-CoV-2.

METHODS

Study Design

The Covid-19 Citizen Science Study is a mobile application that enables the longitudinal and time-updated collection of health survey and location data from thousands of global participants. The application was developed by a team of investigators at the University of California, San Francisco (UCSF) using the Eureka Digital Research Platform. Enrollment began March 26, 2020 and is ongoing (https://covid19.eurekaplatform.org/). The current analysis included participant information collected until October 8, 2020. Enrollment is available to all adults over the age of 18 years and has been facilitated by press releases, social media, and word-of-mouth.

Ethics Approval

Informed electronic consent was obtained remotely using the mobile application at time of study enrollment. The study was approved by the University of California, San Francisco Institutional Review Board (#17-21879).

Data Collection

Surveys collected information about demographics, medical comorbidities, SARS-CoV-2 infection status, daily behaviors, environmental or social exposures, and symptoms. Surveys were written in lay language and met the Flesch-Kincaid criteria for 8th grade reading level (https://readabilityformulas.com). Participants received a baseline survey at time of enrollment ascertaining general demographic information such as age, race/ethnicity, sex, education level, MacArthur subjective social status, occupation, the presence of children or pets at home, and preexisting medical comorbidities. After completing the baseline survey, participants then received: daily surveys that inquired about current symptoms, household contacts, and non-household contacts; weekly surveys that assessed changes to individual-level behaviors such as sleep, exercise, smoking patterns, social distancing efforts, hand hygiene, and use of face masks while out in public; and monthly surveys that collected information regarding employment, mood, and alcohol consumption.

The MacArthur subjective social status ladder was used as a previously validated single-item question to capture socioeconomic status of study participants, with higher point ratings indicating higher subjective social status.[15,16] Occupation was dichotomized based on working in healthcare or not. Exercise was defined as self-reported physical activity lasting for at least 20 minutes and resulted in heavy breathing or "break[ing] a sweat," and was categorized into never/rarely, <1 time/month, <1 time/week, approximately weekly, 2-4 days/week, and >4 days/week. Alcohol use was categorized into none, >0 to 7 standard drinks per week, >7 to 14 standard drinks per week, and >14 standard drinks per week. Smoking activity was differentiated by use of cigarettes, e-cigarettes, or marijuana, and then dichotomized by any use in the last 30 days or not. Daily contacts were defined as any non-household individual with which the participant was within 6 feet of during the course of the day.

Participants were queried regarding polymerase chain reaction (PCR) or antigen testing at baseline and during the weekly survey. Using triggered logic, related questions distinguished between evidence of active infection with the PCR test from other tests, such as antibody tests (the latter were not considered sufficient to constitute incident infection). All participants who reported a positive PCR or antibody test for SARS-CoV-2 prior to enrolling in the study were excluded from this analysis. Self-reported positive PCR tests for SARS-CoV-2 were validated by contacting a sample of participants and obtaining documentation of test results (**Supplementary Appendix**).

Patient and Public Involvement

The Covid-19 Citizen Science Study, which remains open to any interested adult with a smartphone, was designed to answer questions most relevant to patients and the lay public, with an emphasis on identifying clinically relevant behaviors and exposures that can be modified or influenced by any individual. The study was launched using the NIH-supported Eureka Digital Research Platform, which was heavily influenced by prior work designing and implementing the Health eHeart Study[17]—from the beginning, these studies have included patients as key stakeholders, such as the Patient Centered Outcome Research Institute-supported Health eHeart Alliance,[18] to assure that the user experience was relatable and understandable to interested participants around the world. Modifications of questions and the basic content of some research questions was derived from participant feedback received ad hoc and as a result of campaigns to solicit novel research questions from participants for incorporation into the study. All participants in the Covid-19 Citizen Science Study are encouraged to help with recruitment, with regular reminders via text messages, push notifications and newsletters to share the link and/ or "text back" with friends and family members. Results are disseminated back to Covid-19 Citizen Scientists in the form of data visualizations and text shared via newsletters, the study website, and links sent via text message or app-based push notification.

Statistical Analyses

Baseline continuous variables are presented using means and standard deviations (SD) or medians and interquartile ranges (IQR), while categorical variables are presented as frequencies (percentages), and compared between participants who reported incident infection and those remaining infection-free using t-tests for continuous variables and chi-squared tests for categorical variables. Pooled logistic regression models for repeated SARS-CoV-2 test results self-reported on the weekly surveys were used to identify factors, obtained from the baseline and earlier weekly and daily surveys, associated with incident infection. We considered demographics; preexisting medical conditions; behavioral contributors such as mask wearing, hand hygiene, and social distancing efforts; and individual exposures such as number of non-household contacts, large gatherings, and visits to gyms, restaurants, and movie theaters. Exposures from earlier weekly and daily surveys were averaged over measurements obtained 4-21 days prior to the weekly survey providing the SARS-CoV-2 test result. All variables associated with SARS-CoV-2 infection with p-values <0.1 in the pooled logistic regression models adjusting for only a 3-knot restricted cubic spline in calendar date were included in a fully adjusted pooled logistic regression model. In a sensitivity analysis, backward deletion was used to select a more parsimonious pooled logistic regression model retaining covariates with p-values <0.05. These models all used robust standard errors to account for clustering of the repeated weekly SARS-CoV-2 test results by participant. Additionally, recognizing the importance of geographic location, sensitivity analyses restricted to US participants were performed accounting for clustering by county-based Federal Information Processing System (FIPS) and zip codes. All analyses used complete case data. Two-tailed p-values <0.05 were considered statistically significant. All statistical analyses were performed using Stata, version 16 (College Station, TX).

RESULTS

After excluding 628 participants with prevalent SARS-CoV-2 infection, 28,575 individuals without a history of SARS-CoV-2 infection at baseline contributed 2,479,149 participant-days of data to the Covid-19 Citizen Science Study across 99 different countries, including all 50 states in the United States (**Figure 1**). Of these participants, 112 (0.4%) developed a SARS-CoV-2 infection during the study period. Differences in participant demographics, baseline comorbidities, behaviors and exposures between participants who became infected during the study period versus those that did not are displayed in **Table 1**.

After adjusting only for age, sex, race/ethnicity, and calendar date: older age, higher education level, higher subjective social status, and increased alcohol use were associated with lower risk, while working in healthcare, a history of human immunodeficiency virus (HIV), e-cigarette use, less exercise frequency, increased number of recent contacts, attending gatherings with at least 10 people, and visiting movie theaters and restaurants were each associated with a higher risk of incident SARS-CoV-2 infection (**Table 2**). Importantly, pertinent factors that failed to exhibit statistically significant relationships included common medical comorbidities like hypertension, diabetes, coronary artery disease, congestive heart failure, atrial fibrillation, asthma, or chronic obstructive pulmonary disease, as well as hand washing practices and mask wearing frequency. Pooled logistic regression models that incorporated all eligible predictors showed that increased age was associated with lower risk of developing a SARS-CoV-2 infection, whereas increased number of contacts, attending events of at least 10 people, and visits to restaurants were associated with significantly higher risk of later testing positive for SARS-CoV-2 (**Figure 2**). Backward stepwise deletion did not change any of the statically significant relationships (**Supplementary Table 1**). Similarly, the sensitivity analysis using county-based FIPS and zip codes as random effects in US-based data did not meaningfully change the results (**Supplementary Tables 2 and 3**).

DISCUSSION

Among an international cohort free of SARS-CoV-2 at baseline and tracked longitudinally, prospectively, and in a time-updated manner, increased number of daily non-household contacts within 6 feet, events of 10 or more individuals, and restaurant visits each independently predicted a higher risk of developing a SARS-CoV-2. Increased age was associated with a lower risk of subsequently developing a SARS-CoV-2 infection.

As of March 22, 2021, there were over 123 million confirmed cases of SARS-CoV-2 and over 2.7 million SARS-CoV-2-related deaths worldwide.[19] The pandemic has been exacerbated by a recent resurgence of a "second wave" of SARS-CoV-2 cases and confirmation of new strains with potentially increased transmissibility. The pandemic has spurred international efforts to improve testing capabilities,[20] identify therapies to treat the novel coronavirus,[21] and develop vaccines designed to prevent it.[22,23] Even as vaccines from biopharmaceutical companies like Pfizer and Moderna are being delivered, distribution to members of the public has been slow in nearly every country and community, with only countries like Israel, the United Arab Emirates, Chile, and the United Kingdom managing to administer at least 40 vaccine doses per 100 people.[24] Until and if production, distribution, administration, and acceptability of approved vaccines can satisfy the overwhelming need throughout the international community, the identification of preventative health behaviors under an individual's control is crucial to the SARS-CoV-2 public health response.

The Covid-19 Citizen Science Study launched on March 26, 2020 and has been ongoing while recommendations to limit disease transmission continue to evolve at variable rates across the globe. The study has been prospectively collecting data through the initial shelter-in-place recommendations in early 2020 and continues to capture changes in behavioral health patterns as the second spike of SARS-CoV-2 infections surmounts. Our study demonstrated an increased association of SARS-CoV-2 infection in individuals who reported higher numbers of recent contacts. In a similar vein, increased attendance of events of 10 or more people and restaurant visits were associated with increased odds for developing SARS-CoV-2. Given our general understanding of disease transmission for respiratory viruses and recent research characterizing the asymptomatic transmission of SARS-CoV-2,[25,26] these findings are bolstered by biologic plausibility. They add to previous research supporting the use of government mandated physical distancing policies to reduce SARS-CoV-2 incidence[27,28] and demonstrate that behaviors to minimize human-to-human interaction could be effective means to lower one's individual risk of contracting SARS-CoV-2. To our knowledge, this is the first longitudinal cohort to determine that such behaviors among individuals prior to infection actually influence risk.

While the lower risk among older individuals may at first glance appear counter-intuitive, this may be consistent with similar protective behaviors and compliance with social distancing behaviors, especially given data reporting disproportionately higher rates of hospitalization and death in older populations infected with SARS-CoV-2.[29,30] If such phenomena were operative, the fact that we were unable to detect differences in such behaviors (such as significant relationships between hand hygiene or mask-wearing) may be due to collinearity with age and/or suboptimal ascertainment of the actual protective approaches utilized by older individuals. Also contrary to most reports, medical comorbidities thought to increase one's risk of morbidity and mortality from SARS-CoV-2[31,32] such as hypertension, diabetes, congestive heart failure, chronic obstructive pulmonary disease, cancer, and history of prior myocardial infarctions were not retained predictors in our multivariate models, suggesting that prior comorbidities may affect one's response to SARS-CoV-2, but may not play a large role in an individual's risk of contracting SARS-CoV-2.

While previous studies have observed benefits in universal masking at the community level,[33,34] our study did not reveal a clear association between an individual's mask wearing behavior and their risk for SARS-CoV-2 infection. Similarly, self-reported frequency of handwashing did not seem to consistently correlate with SARS-CoV-2 incidence as well. Simple frequencies of mask wearing and hand washing behaviors may be too confounded or measured too imprecisely to observe a consistent trend in our data. These negative results should be interpreted cautiously in the context of the study design and insufficient power may render negative results (or lack of associations) less informative than the statistically significant relationships (positive results) that have been observed thus far (even if in the absence of a longitudinal cohort with time-updated assessments as described here).

Our study has a number of important limitations to note. While focusing on individual-level behaviors mitigated issues involving compliance compared to studies examining state or country-level government mandates, self-report is still a subjective process and still prone to bias based on differing definitions of qualitative words (i.e. “sometimes” versus “most times”). However, health survey data were ascertained prospectively and time-updated daily and weekly to minimize recall bias, and self-report remains likely the most effective method to ascertain individual-level behaviors. As the study required smartphone ownership and use, it is possible that the Covid-19 Citizen Science Study participants represent a more affluent and more technologically savvy population compared to the general population. Though this would limit generalizability instead of internal validity, our diverse recruitment methods were meant to mitigate risks of sampling bias. The distribution of study participants throughout nearly 100 different countries and every state in the US provides fairly unprecedented geographical diversity for a study that also ascertains participant-reported behaviors. There are an innumerable number of behaviors that could have been asked on surveys; we limited our questioning to behaviors previously identified by national and international health organizations and/or those with some biological plausibility as effective means of prevention, such as social distancing, handwashing, and the use of face masks. While PCR testing for SARS-CoV-2 relies on a participant's development of symptoms, index of suspicion, and available access to a testing facility, all factors that may have led to underreporting of all SARS-CoV-2 infections in the study population, the use of these tests to identify SARS-CoV-2 infections ensured that our analyses identified risk factors associated with true infection and increased specificity over traditional methods of symptom reporting. Finally, all data in the Covid-19 Citizen Science Study were collected prospectively as an observational study. While this allows for diverse and rapid sampling of a large population to inform global efforts combating the SARS-CoV-2 pandemic, it remains prone to residual and unmeasured confounding.

In conclusion, the Covid-19 Citizen Science Study, in its prospective and time-updated collection of health data, has identified readily modifiable behaviors that may increase one's individual risk for contracting SARS-CoV-2. Increased number of contacts within 6 feet, events of 10 or more people, and visits to restaurants each independently predicted higher risk of contracting SARS-CoV-2 during the pandemic, while one's demographics, prior medical comorbidities, and adherence to hand washing and face mask wearing were not significant predictors for SARS-CoV-2. During a resurgence of SARS-CoV-2 and continued strain on local governments to balance transmission risk with restrictions on daily life, our study provides community leaders and members of the public with at least 3 modifiable health behaviors within an individual's control that may lower one's personal risk of contracting SARS-CoV-2 during this pandemic.

REFERENCES

1. Gollwitzer A, Martel C, Brady WJ, et al. Partisan differences in physical distancing are linked to health outcomes during the COVID-19 pandemic. *Nat Hum Behav*. 2020;4(11):1186-1197. doi:10.1038/s41562-020-00977-7

2. Painter M, Qiu T. Political Beliefs affect Compliance with COVID-19 Social Distancing Orders. *SSRN Electron J*. 2020. doi:10.2139/ssrn.3569098

3. Barrios J, Hochberg Y. *Risk Perception Through the Lens of Politics in the Time of the COVID-19 Pandemic*. Cambridge, MA; 2020. doi:10.3386/w27008

4. Menni C, Valdes AM, Freidin MB, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat Med*. 2020;26(7):1037-1040. doi:10.1038/s41591-020-0916-2

5. Adorni F, Prinelli F, Bianchi F, et al. Self-reported symptoms of SARS-CoV-2 infection in a non-hospitalized population: results from the large Italian web-based EPICOVID19 cross-sectional survey. (Preprint). *JMIR Public Heal Surveill*. June 2020. doi:10.2196/21866

6. Nguyen LH, Drew DA, Graham MS, et al. Risk of COVID-19 among front-line health-care workers and the general community: a prospective cohort study. *Lancet Public Heal*. 2020;5(9):e475-e483. doi:10.1016/S2468-2667(20)30164-X

7. Salmon Ceron D, Bartier S, Hautefort C, et al. Self-reported loss of smell without nasal obstruction to identify COVID-19. The multicenter Coranosmia cohort study. *J Infect*. July 2020. doi:10.1016/j.jinf.2020.07.005

8. Chu DK, Akl EA, Duda S, et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic

review and meta-analysis. *Lancet*. 2020;395(10242):1973-1987. doi:10.1016/S0140-6736(20)31142-9

9. Greenhalgh T, Schmid MB, Czypionka T, Bassler D, Gruer L. Face masks for the public during the covid-19 crisis. *BMJ*. 2020;369(April):m1435. doi:10.1136/bmj.m1435

10. Avram R, Tison GH, Aschbacher K, et al. Real-world heart rate norms in the Health eHeart study. *npj Digit Med*. 2019;2(1). doi:10.1038/s41746-019-0134-9

11. Christensen MA, Dixit S, Dewland TA, et al. Sleep characteristics that predict atrial fibrillation. *Hear Rhythm*. 2018;15(9):1289-1295. doi:10.1016/j.hrthm.2018.05.008

12. Dixit S, Pletcher MJ, Vittinghoff E, et al. Secondhand smoke and atrial fibrillation: Data from the Health eHeart Study. *Hear Rhythm*. 2016;13(1):3-9. doi:10.1016/j.hrthm.2015.08.004

13. Wang JB, Olgin JE, Nah G, et al. Cigarette and e-cigarette dual use and risk of cardiopulmonary symptoms in the Health eHeart Study. *PLoS One*. 2018;13(7):1-14. doi:10.1371/journal.pone.0198681

14. Tison GH, Avram R, Kuhar P, et al. Worldwide Effect of COVID-19 on Physical Activity: A Descriptive Study. *Ann Intern Med*. 2020;173(9):767-770. doi:10.7326/M20-2665

15. Goodman E, Adler NE, Kawachi I, Frazier a L, Huang B, Colditz G a. Adolescents' Perceptions of Social Status: Development and Evaluation of a New Indicator. *Pediatrics*. 2001;108(2):e31-e31. doi:10.1542/peds.108.2.e31

16. Cooper DC, Milic MS, Mills PJ, Bardwell WA, Ziegler MG, Dimsdale JE. Endothelial Function: The Impact of Objective and Subjective Socioeconomic Status on Flow-Mediated Dilation. *Ann Behav Med*. 2010;39(3):222-231. doi:10.1007/s12160-010-9181-9

17. Olgin JE, Pletcher MJ, Marcus GM. The Health eHeart Study. https://www.health-eheartstudy.org/. Published 2021. Accessed April 19, 2021.

18. Patient-Centered Outcomes Research Institute. The Heart eHeart Alliance. https://www.pcori.org/research-results/2015/health-eheart-alliance. Published 2021. Accessed April 19, 2021.

19. Center for Systems Science and Engineering at Johns Hopkins University. COVID-19 Dashboard. https://coronavirus.jhu.edu/map.html. Published 2020. Accessed March 22, 2021.

20. Tromberg BJ, Schwetz TA, Pérez-Stable EJ, et al. Rapid Scaling Up of Covid-19 Diagnostic Testing in the United States — The NIH RADx Initiative. *N Engl J Med*. 2020;383(11):1071-1077. doi:10.1056/NEJMsr2022263

21. Wang D, Li Z, Liu Y. An overview of the safety, clinical application and antiviral research of the COVID-19 therapeutics. *J Infect Public Health*. July 2020. doi:10.1016/j.jiph.2020.07.004

22. COVID-19 therapies and vaccine landscape. *Nat Mater*. 2020;19(8):809. doi:10.1038/s41563-020-0758-9

23. Kaur SP, Gupta V. COVID-19 Vaccine: A comprehensive status report. *Virus Res*. 2020;288(August):198114. doi:10.1016/j.virusres.2020.198114

24. Our World in Data. Coronavirus (COVID-19) Vaccinations. Our World in Data. https://ourworldindata.org/covid-vaccinations. Published 2020. Accessed March 22, 2021.

25. Mizumoto K, Kagaya K, Zarebski A, Chowell G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Eurosurveillance*. 2020;25(10):1-5. doi:10.2807/1560-

7917.ES.2020.25.10.2000180

26. Furukawa NW, Brooks JT, Sobel J. Evidence Supporting Transmission of Severe Acute Respiratory Syndrome Coronavirus 2 While Presymptomatic or Asymptomatic. *Emerg Infect Dis*. 2020;26(7):E1-E6. doi:10.3201/eid2607.201595

27. Courtemanche C, Garuccio J, Le A, Pinkston J, Yelowitz A. Strong Social Distancing Measures In The United States Reduced The COVID-19 Growth Rate. *Health Aff (Millwood)*. 2020;39(7):1237-1246. doi:10.1377/hlthaff.2020.00608

28. Hsiang S, Allen D, Annan-Phan S, et al. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature*. 2020;584(7820):262-267. doi:10.1038/s41586-020-2404-8

29. Centers for Disease Control and Prevention. COVID-NET. https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html#hospitalizations. Published 2021. Accessed January 16, 2021.

30. National Center for Health Statistics. Daily Updates of Totals by Week and State: Provisional Death Counts for Coronavirus Disease 2019 (COVID-19). https://www.cdc.gov/nchs/nvss/vsrr/COVID19/index.htm. Published 2021. Accessed January 16, 2021.

31. Sanyaolu A, Okorie C, Marinkovic A, et al. Comorbidity and its Impact on Patients with COVID-19. *SN Compr Clin Med*. 2020;2(8):1069-1076. doi:10.1007/s42399-020-00363-4

32. Harrison SL, Fazio-Eynullayeva E, Lane DA, Underhill P, Lip GYH. Comorbidities associated with mortality in 31,461 adults with COVID-19 in the United States: A federated electronic medical record analysis. *PLoS Med*. 2020;17(9):1-11. doi:10.1371/JOURNAL.PMED.1003321

406 33. Lyu W, Wehby GL. Community Use Of Face Masks And COVID-19: Evidence From A
407 Natural Experiment Of State Mandates In The US. *Health Aff*. 2020;39(8):1419-1425.
408 doi:10.1377/hlthaff.2020.00818

409 34. Chernozhukov V, Kasahara H, Schrimpf P. Causal impact of masks, policies, behavior on
410 early covid-19 pandemic in the U.S. *J Econom*. 2020;(xxxx):1-40.
411 doi:10.1016/j.jeconom.2020.09.003

412

413

ACKNOWLEDGEMENTS

This work was supported by IU2CEB021881-01 and 3U2CEB021881-05S1 from the NIH/NIBIB to Drs. Marcus, Olgin, and Pletcher.

AUTHOR CONTRIBUTIONS

G.M.M., J.E.O., M.J.P are the principal investigators for the Covid-19 Citizen Science Study. A.L.L., N.D.P., S.J., V.Y., J.H., R.R., D.W., X.B., C.H., H.E. assisted in data collection. All authors interpreted the data. A.L.L., E.V., and G.M.M. wrote the initial manuscript. A.L.L., E.V., S.J., G.N. made the figures. All authors provided critical comments, edited the manuscript, and approved the final manuscript for its submission.

COMPETING INTERESTS

The authors declare that there are no competing interests.

DATA AVAILABILITY

For participant privacy, data from the Covid-19 Citizen Science Study cannot be stored in a publicly available data repository. All requests for data should be directed to corresponding author G.M.M.

FIGURE LEGENDS

Figure 1. Location of all study participants. The blue shading represents the number of participant-days by county within the US and by nation in the world. The red shading illustrates all participants infected by SARS-CoV-2 during the study period.

Figure 2. Forest plot of all eligible predictors in pooled logistic regression models. Higher scores in the MacArthur Subjective Social Status reflect participants with self-reported higher socioeconomic standing. Large gatherings defined as any gatherings in which 10 or more people were present. The reference group for predictors marked with an asterisk (*) were compared to non-Hispanic whites.

	Participants with Incident SARS-CoV-2 (n=112)	Participants without Incident SARS-CoV-2 (n=28,463)	p-value
Age, median (IQR)	46.0 (37.0-55.5)	44.0 (36.0-55.0)	0.84
Age Category, n (%)			0.71
18-29	12 (10.7)	2,594 (9.2)	
30-39	26 (23.2)	7,832 (27.7)	
40-49	31 (27.7)	7,121 (25.2)	
50-59	27 (24.1)	6,041 (21.3)	
60+	16 (14.3)	4,711 (16.6)	
Female Biological Sex, n (%)	71 (65.1)	18,908 (67.7)	0.79
Race/Ethnicity, n (%)			0.25
White	94 (86.2)	23,814 (85.2)	
Black	2 (1.8)	229 (0.8)	
Hispanic (any race)	9 (8.3)	1,902 (6.8)	
Asian or Pacific Islander	1 (0.9)	1,384 (5.0)	
Other (including multiracial)	3 (2.8)	618 (2.2)	
Highest Level of Education, median, n (%)			<0.001
Less than high school	1 (0.9)	101 (0.4	
High school graduate	8 (7.3)	882 (3.2)	
Some college	24 (22.0)	4,091 (14.7)	
College graduate	40 (36.7)	9,891 (35.4)	
Post-graduate	33 (30.3)	12,690 (45.5)	
Other	3 (2.8)	247 (0.9)	
MacArthur Subjective Social Status Ladder, mean (SD)	6.6 (1.5)	6.9 (1.6)	0.054
Working in healthcare, n (%)	31 (27.7)	5719 (20.1)	0.046
Medical Comorbidities, n (%)			
Atrial fibrillation	2 (1.8)	835 (3.0)	0.47
Anemia	8 (7.2)	2,957 (10.5)	0.26
Asthma	9 (8.0)	2,815 (9.9)	0.50
Coronary artery disease	2 (1.8)	693 (2.4)	0.65
Cancer	5 (4.5)	908 (3.2)	0.45
Congestive heart failure	1 (0.9)	174 (0.6)	0.71
COPD	2 (1.8)	444 (1.6)	0.84
Diabetes	8 (7.1)	1,163 (4.1)	0.11
Hypertension	31 (27.7)	5,675 (20.1)	0.045
HIV	3 (2.7)	108 (0.4)	<0.001
Other immunodeficiency	4 (3.6)	542 (1.9)	0.21
History of heart attack	2 (1.8)	283 (1.0)	0.40
Sleep apnea	13 (11.7)	3,019 (10.8)	0.75
History of stroke	2 (1.8)	355 (1.3)	0.60
Alcohol use			0.10
None	26 (24.8)	6,541 (25.7)	

>0 to 7 drinks per week	60 (57.1)	13,362 (52.6)	
>7 to 14 drinks per week	18 (17.1)	3,764 (14.8)	
>14 drinks per week	1 (1.0)	1,743 (6.9)	
Smoking			
Cigarette use in last 30 days	8 (7.2)	1,421 (5.0)	0.29
E-cigarette use in last 30 days	5 (4.5)	723 (2.6)	0.19
Marijuana use in last 30 days	10 (9.0)	2,650 (9.5)	0.87
Sleep duration, median (IQR)	7.0 (6.0-8.0)	7.0 (6.0-8.0)	
Living with children at home, n (%)	34 (30.4)	8,926 (31.6)	0.78
Living with pets at home, n (%)	73 (65.8)	18,442 (64.9)	0.86
Use of face masks, n (%)			0.29
"Never"	10 (9.3)	1,650 (6.0)	
"Sometimes"	17 (15.7)	3,359 (12.2)	
"Most times"	75 (69.4)	20,591 (74.8)	
"Always"	6 (5.6)	1,910 (6.9)	
Handwashing frequency, n (%)			0.32
<1 time/day	0 (0)	55 (0.2)	
~1 time/day	1 (0.9)	341 (1.2)	
2-4 times/day	16 (14.3)	4,699 (16.5)	
5-10 times/day	47 (42.0)	13,866 (48.7)	
>10 times/day	48 (42.9)	9,502 (33.4)	
Exercise frequency, n (%)			<0.001
Never/Rarely	5 (4.5)	1,591 (5.6)	
<1 time/month	16 (14.3)	2,369 (8.3)	
<1 time/week	23 (20.5)	3,678 (12.9)	
~weekly	12 (10.7)	3,668 (12.9)	
2-4 days/week	30 (26.8)	8,956 (31.5)	
>4 days/week	23 (20.5)	8,107 (28.5)	
Number of contacts in the past 24 hours, mean (SD)	3.8 (6.2)	3.1 (7.3)	0.36
Number of events with 10 or more people in the past week, mean (SD)	3.8 (14.0)	1.9 (9.8)	0.035
Number of gym visits in the past week, mean (SD)	0.4 (3.4)	0.9 (6.6)	0.50
Number of visits to movie theaters in the past week, mean (SD)	0.1 (0.9)	0.1 (1.6)	0.81
Number of visits to restaurants in the past week, mean (SD)	3.4 (9.3)	2.2 (7.7)	0.095

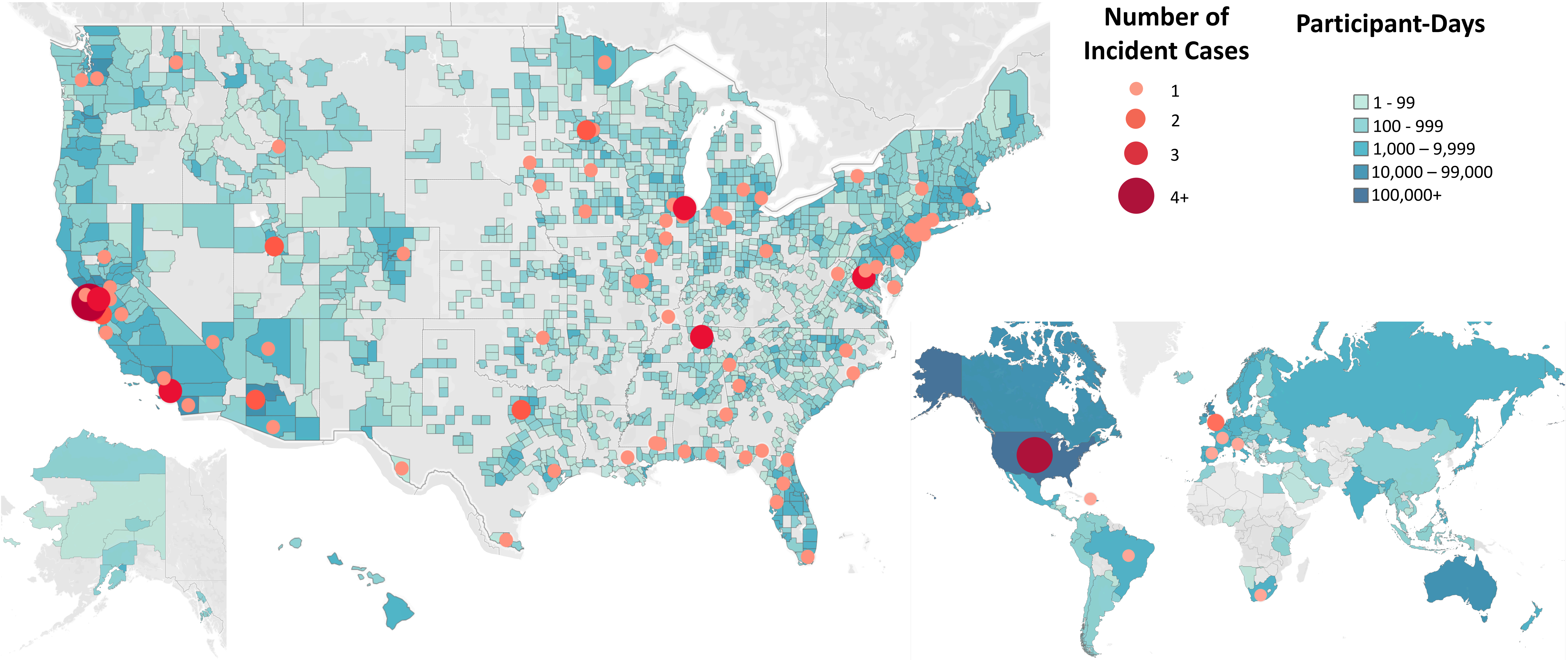
445 **Table 1.** Demographics, comorbidities, and behavioral risk factors of participants in the Covid-
446 19 Citizen Science Study assessed at time of enrollment, divided by participants who later tested
447 positive for Covid-19 during the study period and participants who did not. COPD=chronic

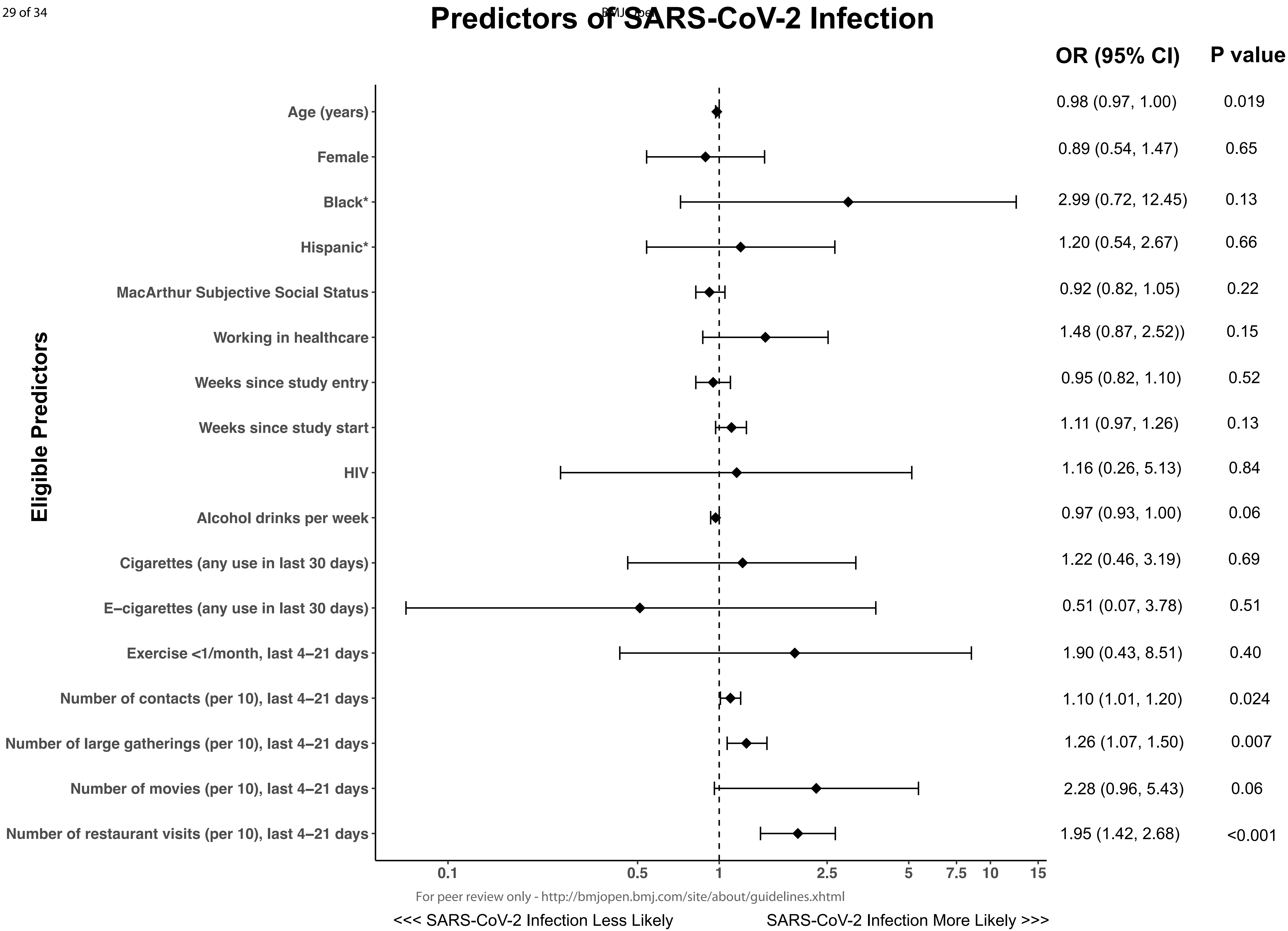
448 obstructive pulmonary disease, HIV=human immunodeficiency virus, IQR=interquartile range,
449 SD=standard deviation.
450
451
452

	Odds Ratio	95% CI	p-value	Group p-value
Age (years)	0.98	0.96, 0.99	<0.001	
Female Biological Sex	0.94	0.63, 1.39	0.76	
Race/Ethnicity				
White	reference			0.35*
Black	2.04	0.50, 8.27	0.32	0.24†
Hispanic (any race)	1.20	0.61, 2.39	0.59	
Asian or Pacific Islander	0.18	0.02, 1.26	0.08	
Other (including multiracial)	1.22	0.39, 3.85	0.73	
Highest Level of Education				
Less than high school	reference			<0.001*
High school graduate	0.91	0.11, 7.44	0.93	<0.001†
Some college	0.51	0.07, 3.87	0.52	0.42#
College graduate	0.34	0.05, 2.56	0.30	
Post-graduate	0.20	0.03, 1.51	0.12	
Other	1.02	0.10, 10.02	0.99	
MacArthur Subjective Social Status Ladder (per point on scale)	0.87	0.79, 0.96	0.004	
Working in healthcare	1.66	1.09, 2.50	0.017	
Medical Comorbidities				
Atrial fibrillation	0.38	0.09, 1.55	0.18	
Anemia	0.65	0.32, 1.34	0.24	
Asthma	0.78	0.40, 1.55	0.48	
Coronary artery disease	0.46	0.11, 1.89	0.28	
Cancer	0.96	0.39, 2.34	0.92	
Congestive heart failure	0.99	0.14, 7.09	0.99	
COPD	0.84	0.21, 3.44	0.81	
Diabetes	1.37	0.67, 2.83	0.39	
Hypertension	1.13	0.75, 1.71	0.56	
HIV	5.31	1.65, 17.12	0.005	
Other immunodeficiency	1.57	0.58, 4.25	0.37	
History of heart attack	1.16	0.28, 4.73	0.84	
Sleep apnea	0.91	0.51, 1.62	0.74	
History of stroke	1.00	0.25, 4.08	1.00	
Alcohol use				
None	reference			0.25*
>0 to 7 drinks per week	0.95	0.60, 1.51	0.83	0.13†
>7 to 14 drinks per week	1.01	0.55, 1.84	0.97	0.047#

>14 drinks per week	0.13	0.02, 0.95	0.044	
Smoking				
Cigarette use in last 30 days	1.91	0.94, 3.88	0.07	
E-cigarette use in last 30 days	2.98	1.64, 5.41	<0.001	
Marijuana use in last 30 days	1.03	0.56, 1.84	0.93	
Mean sleep duration (per hour of sleep)	1.13	0.86, 1.49	0.37	
Living with children at home	1.23	0.89, 1.71	0.21	
Living with pets at home	1.35	0.88, 2.07	0.17	
Use of face masks, last 4-21 days				
"Never"	reference			
"Sometimes"	1.15	0.50, 2.61	0.74	
"Most times" or "Always"	1.11	0.45, 2.72	0.82	
Handwashing frequency, last 4-21 days				
<2 times/day	reference			
2-4 times/day	1.36	0.65, 2.81	0.41	
5-10 times/day	1.08	0.59, 1.95	0.80	
>10 times/day	1.50	0.81, 2.77	0.20	
Exercise frequency, last 4-21 days				
<1 time/month	reference			
<1 time/week	2.21	1.31, 3.76	0.003	
~weekly	1.25	0.76, 2.04	0.38	
2-4 days/week	1.18	0.73, 1.92	0.50	
>4 days/week	0.91	0.51, 1.64	0.76	
Number of contacts (per 10), last 4-21 days	1.17	1.09, 1.26	<0.001	
Number of events with 10 or more people (per 10), last 4-21 days	1.04	1.03, 1.05	<0.001	
Number of gym visits (per 10), last 4-21 days	0.59	0.15, 2.35	0.45	
Number of visits to movie theaters (per 10), last 4-21 days	2.17	1.10, 4.27	0.025	
Number of visits to restaurants (per 10), last 4-21 days	2.06	1.57, 2.70	<0.001	

Table 2. Minimally adjusted odds of incident SARS-CoV-2 infection. Models were adjusted for age, sex, race/ethnicity, and calendar date.
* overall heterogeneity
† heterogeneity of non-reference levels
# linear trend


Number of Incident Cases
1
2
3
4+
Participant-Days
1 - 99
100 - 999
1,000 – 9,999
10,000 – 99,000
100,000+


Predictors of SARS-CoV-2 Infection
OR (95% CI)
P value
Eligible Predictors
Age (years)
0.98 (0.97, 1.00)
0.019
Female
0.89 (0.54, 1.47)
0.65
Black*
2.99 (0.72, 12.45)
0.13
Hispanic*
1.20 (0.54, 2.67)
0.66
MacArthur Subjective Social Status
0.92 (0.82, 1.05)
0.22
Working in healthcare
1.48 (0.87, 2.52))
0.15
Weeks since study entry
0.95 (0.82, 1.10)
0.52
Weeks since study start
1.11 (0.97, 1.26)
0.13
HIV
1.16 (0.26, 5.13)
0.84
Alcohol drinks per week
0.97 (0.93, 1.00)
0.06
Cigarettes (any use in last 30 days)
1.22 (0.46, 3.19)
0.69
E–cigarettes (any use in last 30 days)
0.51 (0.07, 3.78)
0.51
Exercise <1/month, last 4–21 days
1.90 (0.43, 8.51)
0.40
Number of contacts (per 10), last 4–21 days
1.10 (1.01, 1.20)
0.024
Number of large gatherings (per 10), last 4–21 days
1.26 (1.07, 1.50)
0.007
Number of movies (per 10), last 4–21 days
2.28 (0.96, 5.43)
0.06
Number of restaurant visits (per 10), last 4–21 days
1.95 (1.42, 2.68)
<0.001
0.1
0.5
1
2.5
5
7.5
10
15
<<< SARS-CoV-2 Infection Less Likely
SARS-CoV-2 Infection More Likely >>>

SUPPLEMENT

Contents	page

Supplementary Appendix

Participants of the Covid-19 Citizen Science Study who reported a positive polymerase chain reaction (PCR), antigen, or antibody test prior to enrollment in the study or during their time in the study were called by clinical research coordinators to verify their results and request test documentation to be sent to the study coordinators. In a similar manner to participation in the study, submission of test documentation was entirely voluntary. Thus far, 200 participants who reported prevalent or incident SARS-CoV-2 infections have been called to verify their self-reported results. Of the 93 participants who were reached, 83 verbalized that they would send in their test results, and we have thus far received 52 pieces of documentation to verify self-reported SARS-CoV-2 results. Of the 52 pieces of documentation received, all 52 were either laboratory test results or mandated reporting letters from hospitals/clinics notifying the participant of their PCR or antigen-confirmed SARS-CoV-2 infection.

Supplementary Tables

	Odds Ratio	95% CI	p-value	Group p-value
Age (years)	0.98	0.97, 1.00	0.014	
Female Biological Sex	0.95	0.59, 1.54	0.84	
Race/Ethnicity				

White	reference			0.40*
Black	2.96	0.71, 12.29	0.13	0.52†
Hispanic (any race)	1.19	0.53, 2.65	0.67	
Other (including multiracial)	1.69	0.53, 5.40	0.38	
MacArthur Subjective Social Status Ladder	0.92	0.82, 1.04	0.19	
Alcoholic drinks per week, last 4-21 days	0.97	0.93, 1.00	0.07	
Number of contacts (per 10), last 4-21 days	1.11	1.02, 1.21	0.012	
Number of events with 10 or more people (per 10), last 4-21 days	1.26	1.07, 1.48	0.006	
Number of visits to movie theaters (per 10), last 4-21 days	2.00	0.97, 4.11	0.06	
Number of visits to restaurants (per 10), last 4-21 days	1.85	1.37, 2.49	<0.001	
Weeks since study start (linear)	1.04	1.01, 1.07	0.017	

Supplementary Table 1. Backward stepwise logistic model for incident SARS-CoV-2 infection using retention criterion of $p<0.1$ with standard errors clustered on participants.
* overall heterogeneity
† heterogeneity of non-reference levels
# linear trend

	Odds Ratio	**95% CI**	**p-value**	**Group p-value**
Age (years)	0.98	0.96, 0.99	0.008	
Female Biological Sex	0.81	0.49, 1.34	0.42	
Race/Ethnicity				
White	reference			0.43*
Black	3.00	0.72, 12.53	0.13	0.56†
Hispanic (any race)	1.35	0.64, 2.86	0.43	
Other (including multiracial)	1.19	0.28, 4.97	0.81	
MacArthur Subjective Social Status Ladder	0.93	0.82, 1.05	0.24	
Alcoholic drinks per week, last 4-21 days	0.97	0.94, 1.00	0.06	
Number of contacts (per 10), last 4-21 days	1.10	1.00, 1.21	0.04	

Number of events with 10 or more people (per 10), last 4-21 days	1.29	1.09, 1.53	0.003	
Number of visits to movie theaters (per 10), last 4-21 days	1.99	0.97, 4.08	0.059	
Number of visits to restaurants (per 10), last 4-21 days	2.31	1.46, 3.63	<0.001	
Weeks since study start (linear)	1.04	1.01, 1.07	0.008	

Supplementary Table 2. Backward stepwise logistic model for incident SARS-CoV-2 infection using retention criterion of p<0.1 with standard errors clustered on FIPS county-level codes (using US participants only).

* overall heterogeneity

† heterogeneity of non-reference levels

linear trend

	Odds Ratio	**95% CI**	**p-value**	**Group p-value**
Age (years)	0.98	0.97, 0.99	0.007	
Female Biological Sex	0.88	0.55, 1.42	0.60	
Race/Ethnicity				
White	reference			0.39*
Black	2.91	0.70, 12.09	0.14	0.55†
Hispanic (any race)	1.23	0.55, 2.75	0.62	
Other (including multiracial)	1.74	0.54, 5.57	0.35	
MacArthur Subjective Social Status Ladder	0.92	0.81, 1.04	0.17	
Alcoholic drinks per week, last 4-21 days	0.97	0.93, 1.00	0.07	
Number of contacts (per 10), last 4-21 days	1.12	1.03, 1.21	0.008	
Number of events with 10 or more people (per 10), last 4-21 days	1.29	1.09, 1.52	0.003	
Number of visits to movie theaters (per 10), last 4-21 days	1.98	0.95, 4.09	0.07	
Number of visits to restaurants (per 10), last 4-21 days	1.83	1.36, 2.47	<0.001	
Weeks since study start (linear)	1.04	1.01, 1.07	0.015	

Supplementary Table 3. Backward stepwise logistic model for incident SARS-CoV-2 infection using retention criterion of $p<0.1$ with standard errors clustered on zip codes (using US participants only).

* overall heterogeneity

† heterogeneity of non-reference levels

linear trend

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract (*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	5-6
Objectives	3	State specific objectives, including any prespecified hypotheses	6
Methods			
Study design	4	Present key elements of study design early in the paper	6-8
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6-8
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up (*b*) For matched studies, give matching criteria and number of exposed and unexposed	7
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	7-8
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	7-8
Bias	9	Describe any efforts to address potential sources of bias	6-7
Study size	10	Explain how the study size was arrived at	6-7
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	8-9
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding (*b*) Describe any methods used to examine subgroups and interactions (*c*) Explain how missing data were addressed (*d*) If applicable, explain how loss to follow-up was addressed (*e*) Describe any sensitivity analyses	8-9
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed (b) Give reasons for non-participation at each stage (c) Consider use of a flow diagram	9-10
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders (b) Indicate number of participants with missing data for each variable of interest (c) Summarise follow-up time (eg, average and total amount)	9-10
Outcome data	15*	Report numbers of outcome events or summary measures over time	9-10

Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included (*b*) Report category boundaries when continuous variables were categorized (*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	9-10
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	10
Discussion			
Key results	18	Summarise key results with reference to study objectives	10-11
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	13-14
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	14
Generalisability	21	Discuss the generalisability (external validity) of the study results	13-14
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	20

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.

BMJ Open

Predictors of Incident SARS-CoV-2 Infections in an International Prospective Cohort Study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-052025.R1
Article Type:	Original research
Date Submitted by the Author:	06-Aug-2021
Complete List of Authors:	Lin, Anthony; University of California San Francisco, Vittinghoff, Eric; University of California San Francisco, Epidemiology and Biostatistics Olgin, J; University of California San Francisco Peyser, Noah; University of California San Francisco Aung, Sidney; University of California San Francisco Joyce, Sean; University of California San Francisco Yang, Vivian; University of California San Francisco Hwang, Janet; University of California San Francisco Avram, Robert; University of California San Francisco Nah, Gregory; University of California San Francisco Tison, Geoffrey ; University of California San Francisco, Beatty, Alexis; University of California San Francisco, Runge, Ryan; University of California San Francisco Wen, David; University of California San Francisco Butcher, Xochitl; University of California San Francisco Horner, Cathy; University of California San Francisco Eitel, Helena; University of California San Francisco Pletcher, Mark; University of California San Francisco Marcus, GM; University of California San Francisco
Primary Subject Heading:	Public health
Secondary Subject Heading:	Infectious diseases, Health informatics
Keywords:	COVID-19, PUBLIC HEALTH, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, Public health < INFECTIOUS DISEASES

SCHOLARONE™
Manuscripts

Predictors of Incident SARS-CoV-2 Infections in an International Prospective Cohort Study

Anthony L Lin, MD,[1] Eric Vittinghoff, PhD,[2] Jeffrey E Olgin, MD,[3] Noah D Peyser, PhD,[3] Sidney Aung,[4] Sean Joyce,[3] Vivian Yang,[3] Janet Hwang,[3] Robert Avram, MD,[3] Gregory Nah, MA,[3] Geoffrey H Tison, MD, MPH,[3] Alexis L Beatty, MD, MAS,[3] Ryan Runge,[3] David Wen,[3] Xochitl Butcher,[3] Cathy Horner,[3] Helena Eitel,[3] Mark J Pletcher, MD, MPH,[2] Gregory M Marcus, MD, MAS[3]

Author Affiliations

[1]Department of Medicine, University of California, San Francisco

[2]Department of Epidemiology and Biostatistics, University of California, San Francisco

[3]Division of Cardiology, Department of Medicine, University of California, San Francisco

[4]School of Medicine, University of California, San Francisco

Corresponding Author

Gregory M Marcus, MD, MAS

505 Parnassus Ave, M1180B

San Francisco, CA 94143

Phone: 415-476-5706

Fax: 415-476-6260

Email: greg.marcus@ucsf.edu

Word Count: 3,054

ABSTRACT

Objective: Until effective treatments and vaccines are made readily and widely available, preventative behavioral health measures will be central to the SARS-CoV-2 public health response. While current recommendations are grounded in general infectious disease prevention practices, it is still not entirely understood which particular behaviors or exposures meaningfully affect one's own risk of incident SARS-CoV-2 infection. Our objective is to identify individual-level factors associated with one's personal risk for contracting SARS-CoV-2.

Design: Prospective cohort study of adult participants from March 26, 2020 to October 8, 2020.

Setting: The Covid-19 Citizen Science Study, an international, community and mobile-based study collecting daily, weekly, and monthly surveys in a prospective and time-updated manner.

Participants: All adult participants over the age of 18 years were eligible for enrollment.

Primary Outcome Measure: The primary outcome was incident SARS-CoV-2 infection confirmed via polymerase chain reaction or antigen testing.

Results: 28,575 unique participants contributed 2,479,149 participant-days of data across 99 different countries. Of these participants without a history of SARS-CoV-2 infection at time of enrollment, 112 developed an incident infection. Pooled logistic regression models showed that increased age was associated with lower risk (OR 0.98 per year, 95% CI 0.97-1.00, p=0.019), whereas increased number of non-household contacts (OR 1.10 per 10 contacts, 95% CI 1.01-1.20, p=0.024), attending events of at least 10 people (OR 1.26 per 10 events, 95% CI 1.07-1.50, p=0.007), and restaurant visits (OR 1.95 per 10 visits, 95% CI 1.42-2.68, p<0.001) were associated with significantly higher risk of incident SARS-CoV-2 infection.

Conclusions: Our study identified 3 modifiable health behaviors, including the number of non-household contacts, attending large gatherings, and restaurant visits that may meaningfully influence individual-level risk of contracting SARS-CoV-2.

Strengths and limitations of this study

- This large international cohort study with 2.4 million participant-days of data from participants in 99 different countries provides unprecedented geographical diversity for a study analyzing individual-level factors associated with SARS-CoV-2 risk.
- All participants included in this study were free of SARS-CoV-2 infection early in the pandemic, allowing for the real-time ascertainment of significant individual-level behaviors and exposures related to higher risk of incident infection.
- Using polymerase chain reaction or antigen testing as the gold standard for SARS-CoV-2 infections relied on a participant's development of symptoms, index of suspicion, and access to testing facilities, but ensured our study identified risk factors associated with true infection and increased specificity over traditional methods of symptom reporting.

INTRODUCTION

The novel coronavirus (SARS-CoV-2) global pandemic has created a major public health crisis for nearly every country and community in the world. Responses to mitigate transmission have varied by government, but have generally been grounded in known respiratory virus disease prevention practices. Current strategies have included a combination of social distancing, limitations to travel and public gatherings, increased handwashing practices, and the use of face masks. While these interventions are believed to reduce human-to-human transmission, efforts to study these interventions have been limited as they rely on individual-level behaviors that are dynamic with policy changes and can be difficult to capture at scale. Furthermore, the politicization of social distancing recommendations[1–3] makes it difficult to fully understand levels of compliance at the individual-level, and calls for a larger evidence base for recommendations like hand washing, face mask wearing, and limiting human contact, large social gatherings, and visits to restaurants. Identifying predictors of infection requires a longitudinal cohort study. The information gleaned from the longitudinal characterization of SARS-CoV-2 infection risk factors may be crucial to understanding which strategies are most effective and can further inform public policy. Moreover, such data may help elucidate the individual behaviors directly under one’s control to influence one’s personal risk of contracting SARS-CoV-2.

While previous prospective studies have focused primarily on symptom detection and the constellation of symptoms associated with SARS-CoV-2 infection,[4–7] mobile technology provides an opportunity to study the effects of various exposures and behaviors that can be ascertained prospectively, repeatedly, and in nearly real-time. The majority of previous research regarding SARS-CoV-2 has focused on hospitalized individuals, primarily those who already have the disease, and predictors of disease severity as opposed to those pertinent to developing infection. This is not surprising as accumulating sufficient numbers to characterize non-infected individuals at baseline and then follow them over time is generally time-consuming and would require enrollment of particularly large numbers to derive useful results. While systematic reviews and meta-analyses of previous studies have investigated the efficacy of behavioral interventions,[8,9] we are not aware of a longitudinal cohort study in which risk factors have been characterized in detail prior to infection and exposures and behaviors tracked as individuals contracted (or did not contract) SARS-CoV-2 in the community.

Given the widespread use of smartphones and associated mobile apps, the technology is now available to regularly query large populations to assess patterns in SARS-CoV-2 infection rates based on individual-level exposures and behaviors. We have previously demonstrated the utility of this technology in characterizing ambulatory cardiovascular risk factors.[10–14] In this study, we sought to use prospectively-collected information from the Covid-19 Citizen Science Study to identify individual characteristics, exposures, or behaviors associated with an increased risk of contracting SARS-CoV-2.

METHODS

Study Design

The Covid-19 Citizen Science Study is a mobile application that enables the longitudinal and time-updated collection of health survey and location data from thousands of global participants. The application was developed by a team of investigators at the University of California, San Francisco (UCSF) using the Eureka Digital Research Platform. Enrollment began March 26, 2020 and is ongoing (https://covid19.eurekaplatform.org/). The current analysis included participant information collected until October 8, 2020. Enrollment is available to all adults over the age of 18 years and has been facilitated by press releases, social media, and word-of-mouth.

Ethics Approval

Informed electronic consent was obtained remotely using the mobile application at time of study enrollment. The study was approved by the University of California, San Francisco Institutional Review Board (#17-21879).

Data Collection

Surveys collected information about demographics, medical comorbidities, SARS-CoV-2 infection status, daily behaviors, environmental or social exposures, and symptoms. Surveys were written in English and met the Flesch-Kincaid criteria for 8th grade reading level (https://readabilityformulas.com). Participants received a baseline survey at time of enrollment ascertaining general demographic information such as age, race/ethnicity, sex, education level, MacArthur subjective social status, occupation, smoking patterns, the presence of children or pets at home, and preexisting medical comorbidities. After completing the baseline survey, participants then received: daily surveys that inquired about current symptoms, household contacts, and non-household contacts; weekly surveys that assessed changes to individual-level behaviors such as sleep, exercise, social distancing efforts, hand hygiene, and use of face masks while out in public; and monthly surveys that collected information regarding employment, mood, and alcohol consumption (**Supplementary Appendix 1**).

The MacArthur subjective social status ladder was used as a previously validated single-item question to capture socioeconomic status of study participants, with higher point ratings indicating higher subjective social status.[15,16] Occupation was dichotomized based on working in healthcare or not. Exercise was defined as self-reported physical activity lasting for at least 20 minutes and resulted in heavy breathing or “break[ing] a sweat,” and was categorized into never/rarely, <1 time/month, <1 time/week, approximately weekly, 2-4 days/week, and >4 days/week. Alcohol use was categorized into none, >0 to 7 standard drinks per week, >7 to 14 standard drinks per week, and >14 standard drinks per week. Smoking activity was differentiated by use of cigarettes, e-cigarettes, or marijuana, and then dichotomized by any use in the last 30 days or not. Daily contacts were defined as any non-household individual with which the participant was within 6 feet of during the course of the day.

Participants were queried regarding polymerase chain reaction (PCR) or antigen testing at baseline and during the weekly survey. Using triggered logic, related questions distinguished between evidence of active infection with the PCR test from other tests, such as antibody tests (the latter were not considered sufficient to constitute incident infection). All participants who reported a positive PCR or antibody test for SARS-CoV-2 prior to enrolling in the study were excluded from this analysis. Self-reported positive PCR tests for SARS-CoV-2 were validated by contacting a sample of participants and obtaining documentation of test results (**Supplementary Appendix 2**).

Patient and Public Involvement

The Covid-19 Citizen Science Study, which remains open to any interested adult with a smartphone, was designed to answer questions most relevant to patients and the lay public, with an emphasis on identifying clinically relevant behaviors and exposures that can be modified or influenced by any individual. The study was launched using the NIH-supported Eureka Digital Research Platform, which was heavily influenced by prior work designing and implementing the Health eHeart Study[17]—from the beginning, these studies have included patients as key stakeholders, such as the Patient Centered Outcome Research Institute-supported Health eHeart Alliance,[18] to assure that the user experience was relatable and understandable to interested participants around the world. Modifications of questions and the basic content of some research questions was derived from participant feedback received ad hoc and as a result of campaigns to solicit novel research questions from participants for incorporation into the study. All participants in the Covid-19 Citizen Science Study are encouraged to help with recruitment, with regular reminders via text messages, push notifications and newsletters to share the link and/ or "text back" with friends and family members. Results are disseminated back to Covid-19 Citizen Scientists in the form of data visualizations and text shared via newsletters, the study website, and links sent via text message or app-based push notification.

Statistical Analyses

Baseline continuous variables are presented using means and standard deviations (SD) or medians and interquartile ranges (IQR), while categorical variables are presented as frequencies (percentages), and compared between participants who reported incident infection and those remaining infection-free using t-tests for continuous variables and chi-squared tests for categorical variables. Pooled logistic regression models for repeated SARS-CoV-2 test results self-reported on the weekly surveys were used to identify factors, obtained from the baseline and earlier weekly and daily surveys, associated with incident infection. We considered demographics; preexisting medical conditions; behavioral contributors such as mask wearing, hand hygiene, and social distancing efforts; and individual exposures such as number of non-household contacts, large gatherings, and visits to gyms, restaurants, and movie theaters. Exposures from earlier weekly and daily surveys were averaged over measurements obtained 4-21 days prior to the weekly survey providing the SARS-CoV-2 test result. All variables associated with SARS-CoV-2 infection with p-values <0.1 in the pooled logistic regression models adjusting for only a 3-knot restricted cubic spline in calendar date were included in a fully adjusted pooled logistic regression model. In a sensitivity analysis, backward deletion was used to select a more parsimonious pooled logistic regression model retaining covariates with p-values <0.05. These models all used robust standard errors to account for clustering of the repeated weekly SARS-CoV-2 test results by participant. Additionally, recognizing the importance of geographic location, sensitivity analyses restricted to US participants were performed accounting for clustering by county-based Federal Information Processing System (FIPS) and zip codes. All analyses used complete case data. Two-tailed p-values <0.05 were considered statistically significant. All statistical analyses were performed using Stata, version 16 (College Station, TX).

RESULTS

After excluding 628 participants with prevalent SARS-CoV-2 infection, 28,575 individuals without a history of SARS-CoV-2 infection at baseline contributed 2,479,149 participant-days of data to the Covid-19 Citizen Science Study across 99 different countries, including all 50 states in the United States (**Figure 1**). The mean proportion of participants who completed at least one health survey during a study week was 88.6% ± 5.0% and the mean proportion of participants who completed at least one health survey during a study month was 98.1% ± 1.6% (**Supplementary Tables 1 and 2**). Of the total study population, 112 participants (0.4%) developed a SARS-CoV-2 infection during the study period. Differences in participant demographics, baseline comorbidities, behaviors and exposures between participants who became infected during the study period versus those that did not are displayed in **Table 1**.

After adjusting only for age, sex, race/ethnicity, and calendar date: older age, higher education level, higher subjective social status, and increased alcohol use were associated with lower risk, while working in healthcare, a history of human immunodeficiency virus (HIV), e-cigarette use, less exercise frequency, increased number of recent contacts, attending gatherings with at least 10 people, and visiting movie theaters and restaurants were each associated with a higher risk of incident SARS-CoV-2 infection (**Table 2**). Importantly, pertinent factors that failed to exhibit statistically significant relationships included common medical comorbidities like hypertension, diabetes, coronary artery disease, congestive heart failure, atrial fibrillation, asthma, or chronic obstructive pulmonary disease, as well as hand washing practices and mask wearing frequency. Pooled logistic regression models that incorporated all eligible predictors showed that increased age was associated with lower risk of developing a SARS-CoV-2 infection, whereas increased number of contacts, attending events of at least 10 people, and visits to restaurants were associated with significantly higher risk of later testing positive for SARS-CoV-2 (**Figure 2**). Backward stepwise deletion did not change any of the statically significant relationships (**Supplementary Table 3**). Similarly, the sensitivity analysis using county-based FIPS and zip codes as random effects in US-based data did not meaningfully change the results (**Supplementary Tables 4 and 5**).

DISCUSSION

Among an international cohort free of SARS-CoV-2 at baseline and tracked longitudinally, prospectively, and in a time-updated manner, increased number of daily non-household contacts within 6 feet, events of 10 or more individuals, and restaurant visits each independently predicted a higher risk of developing a SARS-CoV-2. Increased age was associated with a lower risk of subsequently developing a SARS-CoV-2 infection.

As of March 22, 2021, there were over 123 million confirmed cases of SARS-CoV-2 and over 2.7 million SARS-CoV-2-related deaths worldwide.[19] The pandemic has been exacerbated by a recent resurgence of a “second wave” of SARS-CoV-2 cases and confirmation of new strains with potentially increased transmissibility. The pandemic has spurred international efforts to improve testing capabilities,[20] identify therapies to treat the novel coronavirus,[21] and develop vaccines designed to prevent it.[22,23] Even as vaccines from biopharmaceutical companies like Pfizer and Moderna are being delivered, distribution to members of the public has been slow in nearly every country and community, with only countries like Israel, the United Arab Emirates, Chile, and the United Kingdom managing to administer at least 40 vaccine doses per 100 people.[24] Until and if production, distribution, administration, and acceptability of approved vaccines can satisfy the overwhelming need throughout the international community, the identification of preventative health behaviors under an individual’s control is crucial to the SARS-CoV-2 public health response.

The Covid-19 Citizen Science Study launched on March 26, 2020 and has been ongoing while recommendations to limit disease transmission continue to evolve at variable rates across the globe. The study has been prospectively collecting data through the initial shelter-in-place recommendations in early 2020 and continues to capture changes in behavioral health patterns as the second spike of SARS-CoV-2 infections surmounts. Our study observed an increased association of SARS-CoV-2 infection in individuals who reported higher numbers of recent contacts. In a similar vein, increased attendance of events of 10 or more people and restaurant visits were associated with increased odds for developing SARS-CoV-2. Given our general understanding of disease transmission for respiratory viruses and recent research characterizing the asymptomatic transmission of SARS-CoV-2,[25,26] these findings are bolstered by biologic plausibility. They add to previous research supporting the use of government mandated physical distancing policies to reduce SARS-CoV-2 incidence[27,28] and suggest that behaviors to minimize human-to-human interaction could be effective means to lower one's individual risk of contracting SARS-CoV-2. To our knowledge, this is the first longitudinal cohort to determine that such behaviors among individuals prior to infection actually influence risk.

While the lower risk among older individuals may at first glance appear counter-intuitive, this may be consistent with similar protective behaviors and compliance with social distancing behaviors, especially given data reporting high incidence of SARS-CoV-2 in nursing homes[29] as well as disproportionately higher rates of hospitalization and death in older populations infected with SARS-CoV-2.[30,31] If such phenomena were operative, the fact that we were unable to detect differences in such behaviors (such as significant relationships between hand hygiene or mask-wearing) may be due to collinearity with age and/or suboptimal ascertainment of the actual protective approaches utilized by older individuals. Also contrary to most reports, medical comorbidities thought to increase one's risk of morbidity and mortality from SARS-CoV-2[32,33] such as hypertension, diabetes, congestive heart failure, chronic obstructive pulmonary disease, cancer, and history of prior myocardial infarctions were not retained predictors in our multivariate models, suggesting that prior comorbidities may affect one's response to SARS-CoV-2, but may not play a large role in an individual's risk of contracting SARS-CoV-2.

While previous studies have observed benefits in universal masking at the community level,[34,35] our study did not reveal a clear association between an individual's mask wearing behavior and their risk for SARS-CoV-2 infection. Similarly, self-reported frequency of handwashing did not seem to consistently correlate with SARS-CoV-2 incidence as well. Simple frequencies of mask wearing and hand washing behaviors may be too confounded or measured too imprecisely to observe a consistent trend in our data. Additionally, the higher prevalence of healthcare workers in the study population may have resulted in participants having higher rates of mask wearing and hand washing, but also higher risk for infection, thereby degrading any associations between predictor and outcome. As such, these negative results should be interpreted cautiously in the context of the study design and insufficient power may render negative results (or lack of associations) less informative than the statistically significant relationships (positive results) that have been observed thus far (even if in the absence of a longitudinal cohort with time-updated assessments as described here).

Our study has a number of important limitations to note. While focusing on individual-level behaviors mitigated issues involving compliance compared to studies examining state or country-level government mandates, self-report is still a subjective process and still prone to bias based on differing definitions of qualitative words (i.e. "sometimes" versus "most times"). However, health survey data were ascertained prospectively and time-updated daily and weekly to minimize recall bias, and self-report remains likely the most effective method to ascertain individual-level behaviors. As the study required smartphone ownership and use, it is possible that the Covid-19 Citizen Science Study participants represent a more affluent and more technologically savvy population compared to the general population. Though this would limit generalizability instead of internal validity, our diverse recruitment methods were meant to mitigate risks of sampling bias. The distribution of study participants throughout nearly 100 different countries and every state in the US provides fairly unprecedented geographical diversity for a study that also ascertains participant-reported behaviors. There are an innumerable number of behaviors that could have been asked on surveys; we limited our questioning to behaviors previously identified by national and international health organizations and/or those with some biological plausibility as effective means of prevention, such as social distancing, handwashing, and the use of face masks. While PCR testing for SARS-CoV-2 relies on a participant's development of symptoms, index of suspicion, and available access to a testing facility, all factors that may have led to underreporting of all SARS-CoV-2 infections in the study population, the use of these tests to identify SARS-CoV-2 infections ensured that our analyses identified risk factors associated with true infection and increased specificity over traditional methods of symptom reporting. Because identification of predictors was determined by testing for statistical significance, we acknowledge that the effect sizes for some of the identified covariates may be small and of questionable clinical relevance. However, this approach enabled us to be as inclusive as possible without constraining potentially relevant predictors based on preconceived assumptions. Finally, all data in the Covid-19 Citizen Science Study were collected prospectively as an observational study. While this allows for diverse and rapid sampling of a large population to inform global efforts combating the SARS-CoV-2 pandemic, it remains prone to residual and unmeasured confounding.

In conclusion, the Covid-19 Citizen Science Study, in its prospective and time-updated collection of health data, has identified readily modifiable behaviors that may increase one's individual risk for contracting SARS-CoV-2. Increased number of contacts within 6 feet, events of 10 or more people, and visits to restaurants each independently predicted higher risk of contracting SARS-CoV-2 during the pandemic, while one's demographics, prior medical comorbidities, and adherence to hand washing and face mask wearing were not significant predictors for SARS-CoV-2. During a resurgence of SARS-CoV-2 and continued strain on local governments to balance transmission risk with restrictions on daily life, our study provides community leaders and members of the public with at least 3 modifiable health behaviors within an individual's control that may lower one's personal risk of contracting SARS-CoV-2 during this pandemic.

REFERENCES

1. Gollwitzer A, Martel C, Brady WJ, et al. Partisan differences in physical distancing are linked to health outcomes during the COVID-19 pandemic. *Nat Hum Behav*. 2020;4(11):1186-1197. doi:10.1038/s41562-020-00977-7

2. Painter M, Qiu T. Political Beliefs affect Compliance with COVID-19 Social Distancing Orders. *SSRN Electron J*. Published online 2020. doi:10.2139/ssrn.3569098

3. Barrios J, Hochberg Y. *Risk Perception Through the Lens of Politics in the Time of the COVID-19 Pandemic*.; 2020. doi:10.3386/w27008

4. Menni C, Valdes AM, Freidin MB, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat Med*. 2020;26(7):1037-1040. doi:10.1038/s41591-020-0916-2

5. Adorni F, Prinelli F, Bianchi F, et al. Self-reported symptoms of SARS-CoV-2 infection in a non-hospitalized population: results from the large Italian web-based EPICOVID19 cross-sectional survey. (Preprint). *JMIR Public Heal Surveill*. Published online June 27, 2020. doi:10.2196/21866

6. Nguyen LH, Drew DA, Graham MS, et al. Risk of COVID-19 among front-line health-care workers and the general community: a prospective cohort study. *Lancet Public Heal*. 2020;5(9):e475-e483. doi:10.1016/S2468-2667(20)30164-X

7. Salmon Ceron D, Bartier S, Hautefort C, et al. Self-reported loss of smell without nasal obstruction to identify COVID-19. The multicenter Coranosmia cohort study. *J Infect*. Published online July 2020. doi:10.1016/j.jinf.2020.07.005

8. Chu DK, Akl EA, Duda S, et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic

review and meta-analysis. *Lancet*. 2020;395(10242):1973-1987. doi:10.1016/S0140-6736(20)31142-9

9. Greenhalgh T, Schmid MB, Czypionka T, Bassler D, Gruer L. Face masks for the public during the covid-19 crisis. *BMJ*. 2020;369(April):m1435. doi:10.1136/bmj.m1435

10. Avram R, Tison GH, Aschbacher K, et al. Real-world heart rate norms in the Health eHeart study. *npj Digit Med*. 2019;2(1). doi:10.1038/s41746-019-0134-9

11. Christensen MA, Dixit S, Dewland TA, et al. Sleep characteristics that predict atrial fibrillation. *Hear Rhythm*. 2018;15(9):1289-1295. doi:10.1016/j.hrthm.2018.05.008

12. Dixit S, Pletcher MJ, Vittinghoff E, et al. Secondhand smoke and atrial fibrillation: Data from the Health eHeart Study. *Hear Rhythm*. 2016;13(1):3-9. doi:10.1016/j.hrthm.2015.08.004

13. Wang JB, Olgin JE, Nah G, et al. Cigarette and e-cigarette dual use and risk of cardiopulmonary symptoms in the Health eHeart Study. *PLoS One*. 2018;13(7):1-14. doi:10.1371/journal.pone.0198681

14. Tison GH, Avram R, Kuhar P, et al. Worldwide Effect of COVID-19 on Physical Activity: A Descriptive Study. *Ann Intern Med*. 2020;173(9):767-770. doi:10.7326/M20-2665

15. Goodman E, Adler NE, Kawachi I, Frazier a L, Huang B, Colditz G a. Adolescents' Perceptions of Social Status: Development and Evaluation of a New Indicator. *Pediatrics*. 2001;108(2):e31-e31. doi:10.1542/peds.108.2.e31

16. Cooper DC, Milic MS, Mills PJ, Bardwell WA, Ziegler MG, Dimsdale JE. Endothelial Function: The Impact of Objective and Subjective Socioeconomic Status on Flow-Mediated Dilation. *Ann Behav Med*. 2010;39(3):222-231. doi:10.1007/s12160-010-9181-9

17. Olgin JE, Pletcher MJ, Marcus GM. The Health eHeart Study. Published 2021. Accessed April 19, 2021. https://www.health-eheartstudy.org/

18. Patient-Centered Outcomes Research Institute. The Heart eHeart Alliance. Published 2021. Accessed April 19, 2021. https://www.pcori.org/research-results/2015/health-eheart-alliance

19. Center for Systems Science and Engineering at Johns Hopkins University. COVID-19 Dashboard. Published 2020. Accessed March 22, 2021. https://coronavirus.jhu.edu/map.html

20. Tromberg BJ, Schwetz TA, Pérez-Stable EJ, et al. Rapid Scaling Up of Covid-19 Diagnostic Testing in the United States — The NIH RADx Initiative. *N Engl J Med*. 2020;383(11):1071-1077. doi:10.1056/NEJMsr2022263

21. Wang D, Li Z, Liu Y. An overview of the safety, clinical application and antiviral research of the COVID-19 therapeutics. *J Infect Public Health*. Published online July 2020. doi:10.1016/j.jiph.2020.07.004

22. COVID-19 therapies and vaccine landscape. *Nat Mater*. 2020;19(8):809. doi:10.1038/s41563-020-0758-9

23. Kaur SP, Gupta V. COVID-19 Vaccine: A comprehensive status report. *Virus Res*. 2020;288(August):198114. doi:10.1016/j.virusres.2020.198114

24. Our World in Data. Coronavirus (COVID-19) Vaccinations. Our World in Data. Published 2020. Accessed March 22, 2021. https://ourworldindata.org/covid-vaccinations

25. Mizumoto K, Kagaya K, Zarebski A, Chowell G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Eurosurveillance*. 2020;25(10):1-5. doi:10.2807/1560-

7917.ES.2020.25.10.2000180

26. Furukawa NW, Brooks JT, Sobel J. Evidence Supporting Transmission of Severe Acute Respiratory Syndrome Coronavirus 2 While Presymptomatic or Asymptomatic. *Emerg Infect Dis*. 2020;26(7):E1-E6. doi:10.3201/eid2607.201595

27. Courtemanche C, Garuccio J, Le A, Pinkston J, Yelowitz A. Strong Social Distancing Measures In The United States Reduced The COVID-19 Growth Rate. *Health Aff (Millwood)*. 2020;39(7):1237-1246. doi:10.1377/hlthaff.2020.00608

28. Hsiang S, Allen D, Annan-Phan S, et al. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature*. 2020;584(7820):262-267. doi:10.1038/s41586-020-2404-8

29. Bagchi S, Mak J, Li Q, et al. Rates of COVID-19 Among Residents and Staff Members in Nursing Homes — United States, May 25–November 22, 2020. *MMWR Morb Mortal Wkly Rep*. 2021;70(2):52-55. doi:10.15585/mmwr.mm7002e2

30. Centers for Disease Control and Prevention. COVID-NET. Published 2021. Accessed January 16, 2021. https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html#hospitalizations

31. National Center for Health Statistics. Daily Updates of Totals by Week and State: Provisional Death Counts for Coronavirus Disease 2019 (COVID-19). Published 2021. Accessed January 16, 2021. https://www.cdc.gov/nchs/nvss/vsrr/COVID19/index.htm

32. Sanyaolu A, Okorie C, Marinkovic A, et al. Comorbidity and its Impact on Patients with COVID-19. *SN Compr Clin Med*. 2020;2(8):1069-1076. doi:10.1007/s42399-020-00363-4

33. Harrison SL, Fazio-Eynullayeva E, Lane DA, Underhill P, Lip GYH. Comorbidities associated with mortality in 31,461 adults with COVID-19 in the United States: A

federated electronic medical record analysis. *PLoS Med*. 2020;17(9):1-11. doi:10.1371/JOURNAL.PMED.1003321

34. Lyu W, Wehby GL. Community Use Of Face Masks And COVID-19: Evidence From A Natural Experiment Of State Mandates In The US. *Health Aff*. 2020;39(8):1419-1425. doi:10.1377/hlthaff.2020.00818

35. Chernozhukov V, Kasahara H, Schrimpf P. Causal impact of masks, policies, behavior on early covid-19 pandemic in the U.S. *J Econom*. 2020;(xxxx):1-40. doi:10.1016/j.jeconom.2020.09.003

FUNDING

This work was supported by IU2CEB021881-01 and 3U2CEB021881-05S1 from the NIH/NIBIB to GMM, JEO, and MJP.

AUTHOR CONTRIBUTIONS

GMM, JEO, and MJP are the principal investigators for the Covid-19 Citizen Science Study and obtained funding for the study. The Covid-19 Citizen Science Study software platform was developed and maintained by NDP, SJ, VY, JH, RR, DW, XB, CH, and HE. ALL, NDP, SJ, VY, JH, RR, DW, XB, CH, and HE assisted in data collection. ALL, EV, JEO, NDP, SA, SJ, VY, JH, RA, GN, GHT, ALB, RR, DW, XB, CH, HE, MJP, and GMM interpreted the data. ALL, EV, and GMM wrote the initial manuscript. ALL, EV, SJ, and GN made the figures. ALL, EV, JEO, NDP, SA, SJ, VY, JH, RA, GN, GHT, ALB, RR, DW, XB, CH, HE, MJP, and GMM provided critical comments during analysis of the data, revised the manuscript, and approved the final manuscript for its submission.

COMPETING INTERESTS

The authors declare that there are no competing interests.

DATA AVAILABILITY

For participant privacy, data from the Covid-19 Citizen Science Study cannot be stored in a publicly available data repository. All requests for data should be directed to corresponding author GMM.

FIGURE LEGENDS

Figure 1. Location of all study participants. The blue shading represents the number of participant-days by county within the US and by nation in the world. The red shading illustrates all participants infected by SARS-CoV-2 during the study period.

Figure 2. Forest plot of all eligible predictors in pooled logistic regression models. Higher scores in the MacArthur Subjective Social Status reflect participants with self-reported higher socioeconomic standing. Large gatherings defined as any gatherings in which 10 or more people were present. The reference group for predictors marked with an asterisk (*) were compared to non-Hispanic whites.

	Participants with Incident SARS-CoV-2 (n=112)	Participants without Incident SARS-CoV-2 (n=28,463)	p-value
Age, median (IQR)	46.0 (37.0-55.5)	44.0 (36.0-55.0)	0.84
Age Category, n (%)			0.71
18-29	12 (10.7)	2,594 (9.2)	
30-39	26 (23.2)	7,832 (27.7)	
40-49	31 (27.7)	7,121 (25.2)	
50-59	27 (24.1)	6,041 (21.3)	
60+	16 (14.3)	4,711 (16.6)	
Female Biological Sex, n (%)	71 (65.1)	18,908 (67.7)	0.79
Race/Ethnicity, n (%)			0.25
White	94 (86.2)	23,814 (85.2)	
Black	2 (1.8)	229 (0.8)	
Hispanic (any race)	9 (8.3)	1,902 (6.8)	
Asian or Pacific Islander	1 (0.9)	1,384 (5.0)	
Other (including multiracial)	3 (2.8)	618 (2.2)	
Highest Level of Education, median, n (%)			<0.001
Less than high school	1 (0.9)	101 (0.4	
High school graduate	8 (7.3)	882 (3.2)	
Some college	24 (22.0)	4,091 (14.7)	
College graduate	40 (36.7)	9,891 (35.4)	
Post-graduate	33 (30.3)	12,690 (45.5)	
Other	3 (2.8)	247 (0.9)	
MacArthur Subjective Social Status Ladder, mean (SD)	6.6 (1.5)	6.9 (1.6)	0.054
Working in healthcare, n (%)	31 (27.7)	5719 (20.1)	0.046
Medical Comorbidities, n (%)			
Atrial fibrillation	2 (1.8)	835 (3.0)	0.47
Anemia	8 (7.2)	2,957 (10.5)	0.26
Asthma	9 (8.0)	2,815 (9.9)	0.50
Coronary artery disease	2 (1.8)	693 (2.4)	0.65
Cancer	5 (4.5)	908 (3.2)	0.45
Congestive heart failure	1 (0.9)	174 (0.6)	0.71
COPD	2 (1.8)	444 (1.6)	0.84
Diabetes	8 (7.1)	1,163 (4.1)	0.11
Hypertension	31 (27.7)	5,675 (20.1)	0.045
HIV	3 (2.7)	108 (0.4)	<0.001
Other immunodeficiency	4 (3.6)	542 (1.9)	0.21
History of heart attack	2 (1.8)	283 (1.0)	0.40
Sleep apnea	13 (11.7)	3,019 (10.8)	0.75
History of stroke	2 (1.8)	355 (1.3)	0.60
Alcohol use			0.10
None	26 (24.8)	6,541 (25.7)	

>0 to 7 drinks per week	60 (57.1)	13,362 (52.6)	
>7 to 14 drinks per week	18 (17.1)	3,764 (14.8)	
>14 drinks per week	1 (1.0)	1,743 (6.9)	
Smoking			
Cigarette use in last 30 days	8 (7.2)	1,421 (5.0)	0.29
E-cigarette use in last 30 days	5 (4.5)	723 (2.6)	0.19
Marijuana use in last 30 days	10 (9.0)	2,650 (9.5)	0.87
Sleep duration, median (IQR)	7.0 (6.0-8.0)	7.0 (6.0-8.0)	
Living with children at home, n (%)	34 (30.4)	8,926 (31.6)	0.78
Living with pets at home, n (%)	73 (65.8)	18,442 (64.9)	0.86
Use of face masks, n (%)			0.29
"Never"	10 (9.3)	1,650 (6.0)	
"Sometimes"	17 (15.7)	3,359 (12.2)	
"Most times"	75 (69.4)	20,591 (74.8)	
"Always"	6 (5.6)	1,910 (6.9)	
Handwashing frequency, n (%)			0.32
<1 time/day	0 (0)	55 (0.2)	
~1 time/day	1 (0.9)	341 (1.2)	
2-4 times/day	16 (14.3)	4,699 (16.5)	
5-10 times/day	47 (42.0)	13,866 (48.7)	
>10 times/day	48 (42.9)	9,502 (33.4)	
Exercise frequency, n (%)			<0.001
Never/Rarely	5 (4.5)	1,591 (5.6)	
<1 time/month	16 (14.3)	2,369 (8.3)	
<1 time/week	23 (20.5)	3,678 (12.9)	
~weekly	12 (10.7)	3,668 (12.9)	
2-4 days/week	30 (26.8)	8,956 (31.5)	
>4 days/week	23 (20.5)	8,107 (28.5)	
Number of contacts in the past 24 hours, mean (SD)	3.8 (6.2)	3.1 (7.3)	0.36
Number of events with 10 or more people in the past week, mean (SD)	3.8 (14.0)	1.9 (9.8)	0.035
Number of gym visits in the past week, mean (SD)	0.4 (3.4)	0.9 (6.6)	0.50
Number of visits to movie theaters in the past week, mean (SD)	0.1 (0.9)	0.1 (1.6)	0.81
Number of visits to restaurants in the past week, mean (SD)	3.4 (9.3)	2.2 (7.7)	0.095

Table 1. Demographics, comorbidities, and behavioral risk factors of participants in the Covid-19 Citizen Science Study assessed at time of enrollment, divided by participants who later tested positive for Covid-19 during the study period and participants who did not. COPD=chronic

obstructive pulmonary disease, HIV=human immunodeficiency virus, IQR=interquartile range, SD=standard deviation.

	Odds Ratio	95% CI	p-value	Group p-value
Age (years)	0.98	0.96, 0.99	<0.001	
Female Biological Sex	0.94	0.63, 1.39	0.76	
Race/Ethnicity				
White	reference			0.35*
Black	2.04	0.50, 8.27	0.32	0.24†
Hispanic (any race)	1.20	0.61, 2.39	0.59	
Asian or Pacific Islander	0.18	0.02, 1.26	0.08	
Other (including multiracial)	1.22	0.39, 3.85	0.73	
Highest Level of Education				
Less than high school	reference			<0.001*
High school graduate	0.91	0.11, 7.44	0.93	<0.001†
Some college	0.51	0.07, 3.87	0.52	0.42#
College graduate	0.34	0.05, 2.56	0.30	
Post-graduate	0.20	0.03, 1.51	0.12	
Other	1.02	0.10, 10.02	0.99	
MacArthur Subjective Social Status Ladder (per point on scale)	0.87	0.79, 0.96	0.004	
Working in healthcare	1.66	1.09, 2.50	0.017	
Medical Comorbidities				
Atrial fibrillation	0.38	0.09, 1.55	0.18	
Anemia	0.65	0.32, 1.34	0.24	
Asthma	0.78	0.40, 1.55	0.48	
Coronary artery disease	0.46	0.11, 1.89	0.28	
Cancer	0.96	0.39, 2.34	0.92	
Congestive heart failure	0.99	0.14, 7.09	0.99	
COPD	0.84	0.21, 3.44	0.81	
Diabetes	1.37	0.67, 2.83	0.39	
Hypertension	1.13	0.75, 1.71	0.56	
HIV	5.31	1.65, 17.12	0.005	
Other immunodeficiency	1.57	0.58, 4.25	0.37	
History of heart attack	1.16	0.28, 4.73	0.84	
Sleep apnea	0.91	0.51, 1.62	0.74	
History of stroke	1.00	0.25, 4.08	1.00	
Alcohol use				
None	reference			0.25*
>0 to 7 drinks per week	0.95	0.60, 1.51	0.83	0.13†
>7 to 14 drinks per week	1.01	0.55, 1.84	0.97	0.047#

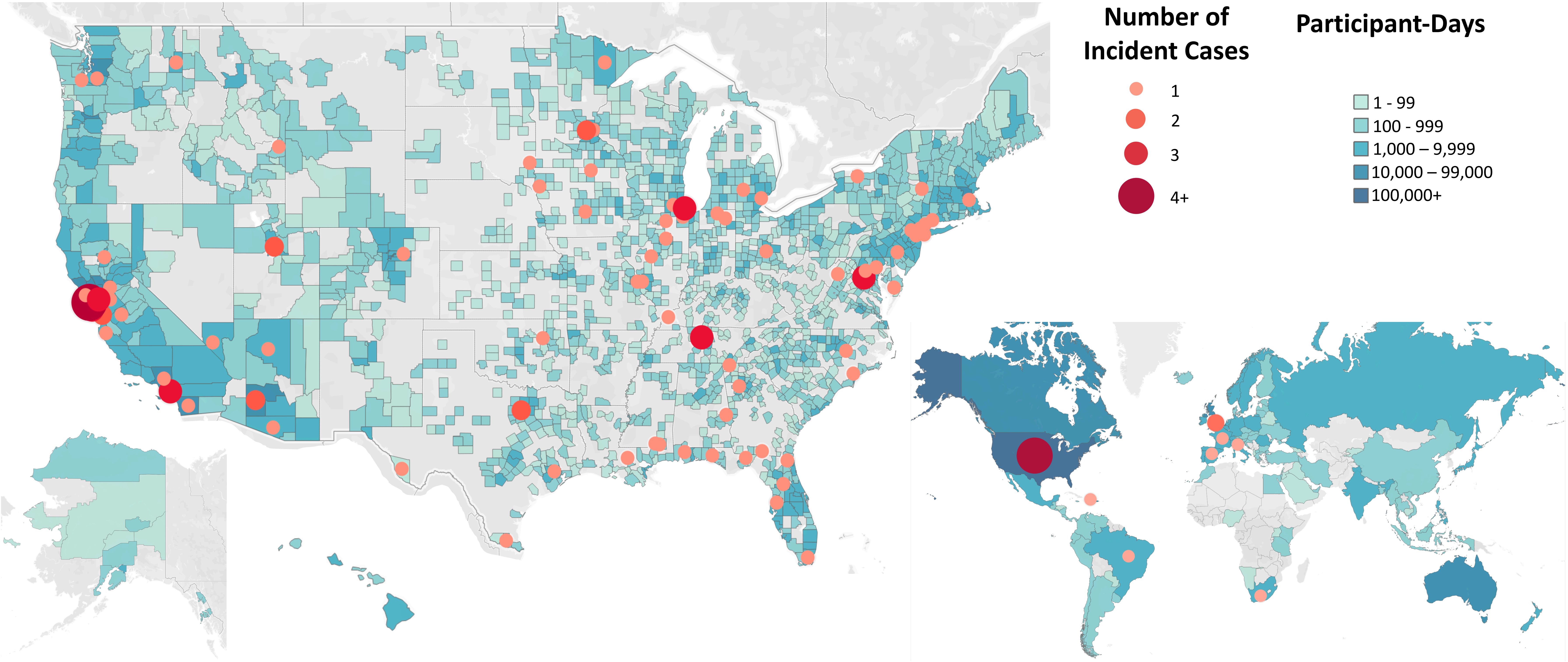
>14 drinks per week	0.13	0.02, 0.95	0.044	
Smoking				
Cigarette use in last 30 days	1.91	0.94, 3.88	0.07	
E-cigarette use in last 30 days	2.98	1.64, 5.41	<0.001	
Marijuana use in last 30 days	1.03	0.56, 1.84	0.93	
Mean sleep duration (per hour of sleep)	1.13	0.86, 1.49	0.37	
Living with children at home	1.23	0.89, 1.71	0.21	
Living with pets at home	1.35	0.88, 2.07	0.17	
Use of face masks, last 4-21 days				
"Never"	reference			
"Sometimes"	1.15	0.50, 2.61	0.74	
"Most times" or "Always"	1.11	0.45, 2.72	0.82	
Handwashing frequency, last 4-21 days				
<2 times/day	reference			
2-4 times/day	1.36	0.65, 2.81	0.41	
5-10 times/day	1.08	0.59, 1.95	0.80	
>10 times/day	1.50	0.81, 2.77	0.20	
Exercise frequency, last 4-21 days				
<1 time/month	reference			
<1 time/week	2.21	1.31, 3.76	0.003	
~weekly	1.25	0.76, 2.04	0.38	
2-4 days/week	1.18	0.73, 1.92	0.50	
>4 days/week	0.91	0.51, 1.64	0.76	
Number of contacts (per 10), last 4-21 days	1.17	1.09, 1.26	<0.001	
Number of events with 10 or more people (per 10), last 4-21 days	1.04	1.03, 1.05	<0.001	
Number of gym visits (per 10), last 4-21 days	0.59	0.15, 2.35	0.45	
Number of visits to movie theaters (per 10), last 4-21 days	2.17	1.10, 4.27	0.025	
Number of visits to restaurants (per 10), last 4-21 days	2.06	1.57, 2.70	<0.001	

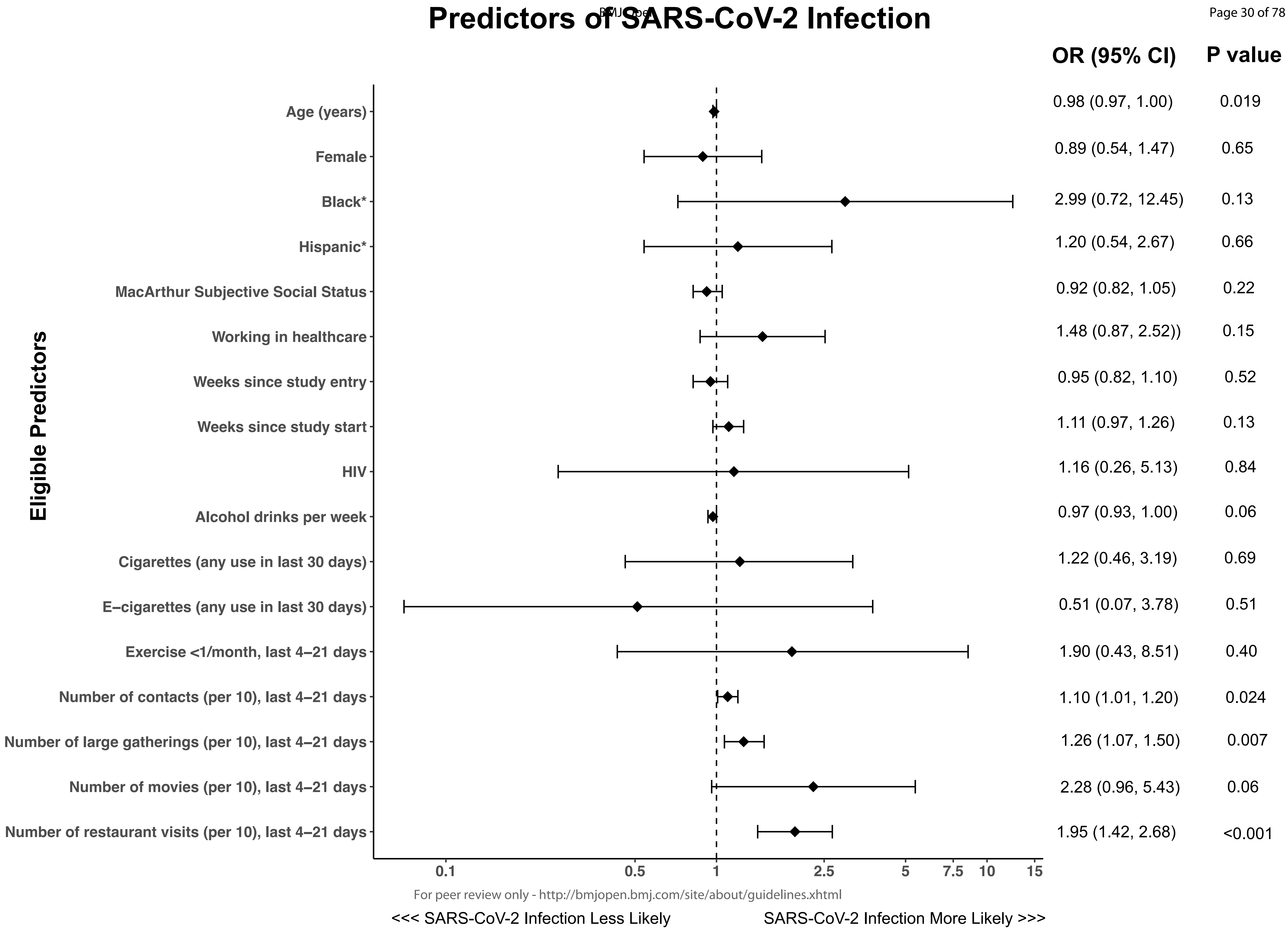
Table 2. Minimally adjusted odds of incident SARS-CoV-2 infection. Models were adjusted for age, sex, race/ethnicity, and calendar date.

* overall heterogeneity

† heterogeneity of non-reference levels

linear trend


Number of Incident Cases
1
2
3
4+
Participant-Days
1 - 99
100 - 999
1,000 – 9,999
10,000 – 99,000
100,000+


Predictors of SARS-CoV-2 Infection
OR (95% CI)
P value
Eligible Predictors
Age (years)
0.98 (0.97, 1.00)
0.019
Female
0.89 (0.54, 1.47)
0.65
Black*
2.99 (0.72, 12.45)
0.13
Hispanic*
1.20 (0.54, 2.67)
0.66
MacArthur Subjective Social Status
0.92 (0.82, 1.05)
0.22
Working in healthcare
1.48 (0.87, 2.52))
0.15
Weeks since study entry
0.95 (0.82, 1.10)
0.52
Weeks since study start
1.11 (0.97, 1.26)
0.13
HIV
1.16 (0.26, 5.13)
0.84
Alcohol drinks per week
0.97 (0.93, 1.00)
0.06
Cigarettes (any use in last 30 days)
1.22 (0.46, 3.19)
0.69
E−cigarettes (any use in last 30 days)
0.51 (0.07, 3.78)
0.51
Exercise <1/month, last 4−21 days
1.90 (0.43, 8.51)
0.40
Number of contacts (per 10), last 4−21 days
1.10 (1.01, 1.20)
0.024
Number of large gatherings (per 10), last 4−21 days
1.26 (1.07, 1.50)
0.007
Number of movies (per 10), last 4−21 days
2.28 (0.96, 5.43)
0.06
Number of restaurant visits (per 10), last 4−21 days
1.95 (1.42, 2.68)
<0.001
0.1
0.5
1
2.5
5
7.5
10
15
<<< SARS-CoV-2 Infection Less Likely
SARS-CoV-2 Infection More Likely >>>

SUPPLEMENT

Contents page

Supplementary Appendix 1

All participants of the Covid-19 Citizen Science Study received health surveys through the mobile application regarding their demographics, medical comorbidities, SARS-CoV-2 infection status, behaviors, and exposures. Baseline surveys were conducted once at time of study enrollment. After enrollment, the mobile application would prompt participants to answer daily, weekly, and monthly health surveys to assess individual behaviors and exposures, as well as inquire about SARS-CoV-2 infection status and associated symptoms. Details regarding the specific questions in the health surveys can be found below.

Sections

Baseline Data Collection

Demographics Survey

Baseline Data Collection

What sex were you assigned at birth?

Male

Female

Prefer not to disclose

How would you describe your current gender identity?

Male

Female

Transgender Woman (Male-to-Female)

Transgender Man (Female-to-Male)

Genderqueer

Another Gender Identity

Decline to state

What gender identity do you identify with? (Optional)

What is your racial background? CHECK ALL THAT APPLY.

- Black or African American
- White
- Asian (including South Asian and Asian Indian)
- Native Hawaiian or Pacific Islander
- American Indian or Alaska Native
- Some other race
- Don't know

What is your Asian background?

- Chinese
- Filipino
- Asian Indian
- Japanese
- Korean
- Vietnamese
- Other Asian or Mix

What is your Pacific Island background?

- Native Hawaiian
- Samoan
- Guamanian or Chamorro
- Other Pacific Islander or Mix

This is a question about ethnicity, rather than race, as used in the US Census. For example, someone may be of white race and Hispanic ethnicity or black race and Hispanic ethnicity. Tap next to continue.

Are you of Hispanic, Latino or Spanish origin or ancestry?

- No
- Yes: Mexican, Mexican American or Chicano
- Yes: Puerto Rican
- Yes: Cuban
- Yes: Other or Mixed Hispanic, Latino or Spanish origin
- Don't know
- Prefer not to state

Think of this ladder as representing where people stand in your country. At the top of the ladder are the people who are the best off -- those who have the most money, the most education and the most respectful jobs. At the bottom are the people who are the worst off -- who have the least money, least education, and least respectful jobs or no job. The higher up you are on the ladder, the closer you are to the people at the very top; the lower you are, the closer you are to the people at the very bottom. Tap next to continue.

Where would you place yourself on this ladder?

What is the highest level of education you have achieved?

- No formal schooling
- Some school, but did not graduate high school
- High school diploma or equivalency (e.g., GED)
- Associate degree (e.g., junior college)
- Some college, but did not graduate college
- Bachelor's degree
- Master's degree
- Doctorate (PhD)
- Professional doctorate (MD, JD, DDS, etc.)
- Other
- Don't know
- Prefer not to state

Click here to finish

Your Medical Conditions

Sections

Baseline Data Collection

Have you ever been told by a doctor or nurse that you have, or have been treated for, any of the following conditions (in the past or currently)? Tap next to continue.

High blood pressure or hypertension (except that occurred during pregnancy and did not last after pregnancy)?

Yes

No

Don't know

Diabetes? Do not include pre-diabetes.

Yes

No

Don't know

Coronary artery disease (blockages in your heart vessels) or angina (chest pain)?

Yes

No

Don't know

A heart attack (myocardial infarction)?

Yes

No

Don't know

Congestive Heart failure (CHF, Heart Failure)?

Yes
No
Don't know

Stroke or TIA (Transient Ischemic Attack or Mini-Stroke)?

Yes
No
Don't know

Atrial fibrillation (Afib, AF)?

Yes
No
Don't know

Sleep apnea (obstructive sleep apnea, OSA)?

Yes
No
Don't know

COPD (emphysema, chronic bronchitis, obstructive pulmonary disease)?

Yes
No
Don't know

Asthma, to the point that you use inhalers daily or have been to the hospital for your asthma?

Yes
No

Don't know

Cancer (including leukemia or lymphoma) undergoing active treatment?

Yes

No

Don't know

Immunodeficiency (NOT including HIV)?

Yes

No

Don't know

Chronic HIV infection?

Yes

No

Don't know

Anemia or other blood disorder (do not include leukemia or lymphoma)?

Yes

No

Don't know

Are you currently pregnant?

Yes

No

Don't know

Sections

Your Smoking History

Baseline Data Collection

Have you ever smoked a cigarette, even one or two puffs?

Yes | No

Don't know | Refuse to answer

Have you smoked cigarettes in the past 30 days?

Yes | No

Refuse to answer

About how many days have you smoked a cigarette in the past 30 days?

On average, how many cigarettes per day have you smoked in the past 30 days (use 1 if less than one)

cigarettes per day

Have you ever smoked a cigar, cigarillo, or tobacco product other than cigarette, even one or two puffs?

Yes | No

Don't know | Refuse to answer

Have you smoked a cigar, cigarillo, or tobacco product other than a cigarette in the past 30 days?

Yes

No

Don't know

Refuse to answer

About how many days have you smoked a cigar, cigarillo, or tobacco product other than cigarette in the past 30 days?

days

On average, how many cigar, cigarillo, or tobacco product (other than cigarettes) per day have you smoked in the past 30 days (use 1 if less than one)?

Have you ever used an electronic nicotine product (e-cigarette, vape nicotine), even one or two puffs?

Yes

No

Don't know

Refuse to answer

Have you used an electronic nicotine product in the past 30 days?

Yes

No

Don't know

Refuse to answer

About how many days did you use it in the past 30 days?

days

How many puffs from an e-cigarette do you typically take over the past 30 days?

How much did you spend on electronic delivery products in the past 30 days?

Dollars

Have you smoked or vaped marijuana, even one or two puffs?

Yes

No

Don't know

Refuse to answer

Have you smoked or vaped marijuana in the past 30 days?

Yes

No

Don't know

Refuse to answer

How many days did you smoke or vape marijuana in the past 30 days?

Days

Sections

Baseline Survey

Baseline Data Collection

In what country is your primary residence?

What is the ZIP code (if in the U.S.) or postal code of your primary residence?

Have you had any of the following symptoms since February 1, 2020 for more than 3 days in a row? CHECK ALL THAT APPLY

- A scratchy throat
- A painful sore throat
- A cough (worse than usual if you have a baseline cough)
- A runny nose
- Symptoms of fever or chills
- A temperature greater than 100.4 °F or 38.0 °C
- Muscle aches (worse than usual if you have baseline muscle aches)
- Nausea, vomiting or diarrhea
- Shortness of breath
- Unable to taste or smell
- Red or painful eyes
- None of the above

Have you ever been tested for the novel coronavirus, the virus that causes COVID-19 (either a test to detect the virus for active infection or the antibody to detect past infection)?

- Yes
- No

Other

Was it a test for active infection (virus) or past infection (antibody to the virus)? (The test for active infection usually uses a swab or saliva; the test for past infection usually uses blood.)

Test for active infection (virus)

Test for past infection (antibody to the virus)

I had both kind of tests

I don't know

Do you think you previously experienced symptomatic infection due to COVID-19?

Yes

No

When did your symptoms start?

What symptoms did you have? CHECK ALL THAT APPLY

A scratchy throat

A painful sore throat

A cough (worse than usual if you have a baseline cough)

A runny nose

Symptoms of fever or chills

A temperature greater than 100.4 °F or 38.0 °C

Muscle aches (worse than usual if you have baseline muscle aches)

Nausea, vomiting or diarrhea

Shortness of breath

Unable to taste or smell

Red or painful eyes

Other

If other, please explain.

During the illness that you believe was due to COVID-19, were you tested for the flu?

Yes

No

What was the result?

Positive for the flu

Negative for the flu

Other

Prior to the illness you believe was due to COVID-19, were you in physical contact with someone else that tested positive for the disease?

Yes

No

Other

Prior to the illness you believe was due to COVID-19, were you in physical contact with someone else with symptoms suggestive of COVID-19?

Yes

No

Other

Prior to the illness you believe was due to COVID-19, had you traveled to a region known to have a high prevalence of COVID-19?

Yes

No

Other

During the illness you believe was due to COVID-19, did you seek to receive a test for active COVID-19 infection?

Yes

No

Other

What happened when you sought the coronavirus test?

I did receive a test, and it was positive.

I did receive a COVID-19 test for active infection, and it was negative.

I did receive a COVID-19 test for

I was evaluated by a healthcare

active infection, but do not know the results.

provider, but they did not believe the test was indicated.

I was evaluated by a healthcare provider and they wanted to order a test, but it was not available.

Other

Do you continue to have symptoms due to the illness you believe to be due to COVID-19?

Yes

No

On what date did you last experience symptoms?

Are there other reasons not covered by this survey that lead you to believe you have been infected with the novel coronavirus?

Yes

No

Other

Please explain.

About how many weeks ago was your test for active COVID-19 infection (virus)? Put 0 if this week.

weeks ago

About how many weeks ago was your test for past infection (antibody to the COVID-19 virus)? Put 0 if this week.

weeks ago

Do you know the result of your test for active COVID-19 infection (virus)?

- Yes, I was positive (the novel coronavirus WAS detected)
- Yes, I was negative (the novel coronavirus was NOT detected)
- Yes, the test was inconclusive
- No, not yet

Do you know the result of your test for past infection (antibody to the COVID-19 virus)?

- Yes, I was positive (antibody to COVID-19 WAS detected suggesting past exposure)
- Yes, I was negative (antibody to COVID-19 was NOT detected suggesting NO past exposure)
- Yes, the test was inconclusive
- No, not yet

Why was the test for active COVID-19 infection (virus) performed? CHECK ALL THAT APPLY

- I had symptoms concerning for COVID-19 infection (including hospitalization for COVID-19)
- I was exposed to someone with suspected or confirmed COVID-19
- Prior to a medical procedure or hospitalization that was unrelated to COVID-19
- It was offered through my healthcare provider as part of routine screening (not related to symptoms or pregnancy)

It was part of screening for my pregnancy

I am a healthcare worker and it is offered or mandated by my employer

As part of a research study

It was required by my work

Part of a public health effort

I obtained it on my own

Not sure or other

Why was the test for past infection (antibody to the COVID-19 virus) performed? CHECK ALL THAT APPLY

I had symptoms concerning for COVID-19 infection (including hospitalization for COVID-19)

I was exposed to someone with suspected or confirmed COVID-19

Prior to a medical procedure or hospitalization that was unrelated to COVID-19

It was offered through my healthcare provider as part of routine screening (not related to symptoms or pregnancy)

It was part of screening for my pregnancy

I am a healthcare worker and it is offered or mandated by my employer

As part of a research study

It was required by my work

Part of a public health effort

I obtained it on my own

Not sure or other

Which of the following describes your primary area of employment?

Healthcare

Education

Retail

Transportation

Arts, entertainment, and recreation

Hospitality and food services

Finance and insurance

Scientific and technical services

Utilities

Construction

Manufacturing

Other

Are you aware of any novel coronavirus (the virus causing COVID-19) infected individuals in your COUNTY (or local area equivalent if your area does not have counties)?

Yes

No

How worried are you that the health of you or your loved ones will be affected by the novel coronavirus (the virus causing COVID-19)?

Extremely worried

Very worried

Somewhat worried

A little worried

Not worried at all

Has your local government issued or continued any of the following restrictions? CHECK ALL THAT APPLY

School closures

Restricted gatherings at (or closed) bars, restaurants, and/or theaters

Restricted gatherings of a certain number of individuals

Recommended working from home or not working

Shelter in place (required to stay home except for essential activities)

Other restrictions

How have your hand hygiene practices (washing hands and/or using hand sanitizer) changed since learning about the novel coronavirus (the virus causing COVID-19)?

I wash or sanitize my hands MUCH MORE frequently than before

I wash or sanitize my hands SOMEWHAT MORE frequently than before

I wash or sanitize my hands A LITTLE MORE frequently than before

I have not made any changes

I wash or sanitize my hands A LITTLE LESS frequently than before

I wash or sanitize my hands SOMEWHAT LESS frequently than before

I wash or sanitize my hands MUCH LESS frequently than before

Have you sanitized your mobile phone (such as by using sanitizing wipes or hand sanitizer) since learning of the novel coronavirus (the virus causing COVID-19)?

Yes

No

Other

Do any school-aged (K-12 or equivalent) children live with you?

Yes

No

Other

Do you have a college-aged child (under the age of 25) who usually does not live in your home but who has returned home and is living in your house because of the coronavirus pandemic?

Yes

No

What date did they return? (Your best guess is fine.)

MM/DD/YYYY

What school were they attending?

School

Where is the school located?

Do you live with or have continued regular in-person contact with an elderly person (over 65 years of age) or someone susceptible to illness (being immunocompromised or having a pre-existing medical condition)?

Yes

No

Other

Do you have any pets at home?

Yes

No

Other

What pets live with you (CHECK ALL THAT APPLY):

Dog(s)

Cat(s)

Bird(s)

Reptile(s)

Other

Did you have a flu shot (influenza vaccine) in the past year?

Yes

No

Other

Have you had cold or flu symptoms (enough that you would say that you had a cold or the flu) in the past year?

Yes

No

How many cold or flu illnesses in the past year were associated with a fever (Temperature > 101.3 F or > 38.5 C)?

None

1-3

4-6

More than 6

When was the last one?

weeks ago

How many cold or flu illnesses in the past year were NOT associated with a fever (Temperature > 101.3 F or > 38.5 C)?

None

1-3

4-6

More than 6

When was the last one?

weeks ago

On average, how often have you exercised (enough to breathe heavily and/or sweat) over the past year?

Never or rarely

Less than once a month

More than once a month but less than once a week

About once a week

More than once a week but less than 4 times a week

4 or more times a week

Other

IN THE PAST WEEK: How many drinks of alcohol (one drink = one standard glass of wine, can of beer, or shot of hard liquor) did you consume?

drinks

Sections

Daily COVID-19 Citizen Science Survey

Daily Surveys

IN THE PAST 24 HOURS, approximately how many people outside of your household did you interact with while they were within 6 feet? ("Interact" is loosely defined as talking, touching, or just being within 6 ft of someone for longer than 1 or 2 minutes).

people

Approximately what percent of those people were wearing masks, or were behind a shield?

%

IN THE PAST 24 HOURS: have YOU had any of the following (CHECK ALL THAT APPLY):

- A scratchy throat
- A painful sore throat
- A cough (worse than usual if you have a baseline cough)
- A runny nose
- Symptoms of fever or chills
- A temperature greater than 100.4 °F or 38.0 °C
- Muscle aches (worse than usual if you have baseline muscle aches)
- Nausea, vomiting or diarrhea
- Shortness of breath
- Unable to taste or smell
- Red or painful eyes
- None of the above

Did you seek medical care for these symptoms?

Yes

No

IN THE PAST 24 HOURS, has ANYONE (other than you) in your household had ANY of those symptoms? (scratchy/sore throat, cough, runny nose, fevers/chills/high temperature, muscle aches, nausea/vomiting/diarrhea, shortness of breath, unable to taste or smell, red or painful eyes)

Yes

No

Not sure

Sections

Weekly COVID-19 Citizen Science Survey

Weekly Surveys

In the past week, have you received results of any tests that you had done for the novel coronavirus, the virus that causes COVID-19 (either a test to detect the virus for active infection or the antibody to detect past infection)?

- Yes
- No
- I got a test, but don't know the results

Was it a test for active infection (virus) or past infection (antibody to the virus)? (The test for active infection usually uses a swab or saliva; the test for past infection usually uses blood.)

- Test for active infection (virus)
- Test for past infection (antibody to the virus)
- I had both kind of tests
- I don't know

Do you know the result of your test for active COVID-19 infection (virus)?

- Yes, I was positive (the novel coronavirus WAS detected)
- Yes, I was negative (the novel coronavirus was NOT detected)
- Yes, the test was inconclusive
- No, not yet

Do you know the result of your test for past infection (antibody to the COVID-19 virus)?

- Yes, I was positive (antibody to
- Yes, I was negative (antibody to

COVID-19 WAS detected suggesting past exposure)

COVID-19 was NOT detected suggesting NO past exposure)

Yes, the test was inconclusive

No, not yet

Why was the test for active COVID-19 infection (virus) performed? CHECK ALL THAT APPLY

I had symptoms concerning for COVID-19 infection (including hospitalization for COVID-19)

I was exposed to someone with suspected or confirmed COVID-19

Prior to a medical procedure or hospitalization that was unrelated to COVID-19

It was offered through my healthcare provider as part of routine screening (not related to symptoms or pregnancy)

It was part of screening for my pregnancy

I am a healthcare worker and it is offered or mandated by my employer

As part of a research study

It was required by my work

Part of a public health effort

I obtained it on my own

Why was the test for past infection (antibody to the COVID-19 virus) performed? CHECK ALL THAT APPLY

I had symptoms concerning for COVID-19 infection (including hospitalization for COVID-19)

I was exposed to someone with suspected or confirmed COVID-19

Prior to a medical procedure or

It was offered through my

hospitalization that was unrelated to COVID-19

healthcare provider as part of routine screening (not related to symptoms or pregnancy)

It was part of screening for my pregnancy

I am a healthcare worker and it is offered or mandated by my employer

As part of a research study

It was required by my work

Part of a public health effort

I obtained it on my own

Over the past WEEK, how worried have you been that the health of you or your loved ones will be affected by the novel coronavirus (the virus causing COVID-19)?

Extremely worried

Very worried

Somewhat worried

A little worried

Not worried at all

Over the past WEEK, on average, how often have you washed or sanitized your hands?

More than 10 times per day

5-10 times per day

2-4 times per day

About once per day

Less than once per day

Over the past WEEK, how many times have you visited a gym?

Over the past WEEK, how many times have you visited a restaurant (not for takeout)?

Over the past WEEK, how many times have you visited a bar?

Over the past WEEK, how many times have you visited a movie theater?

Over the past WEEK, how many times have you visited a grocery store or pharmacy?

Over the past WEEK, how many times have you visited an event with more than 10 people?

Over the past WEEK, how many times have you exercised for more than 20 minutes (enough to breathe heavily and/or sweat)?

Over the past WEEK, has your local government issued or continued any of the following restrictions? (CHECK ALL THAT APPLY)

- Shelter in place (required to stay home except essential activities)
- Wearing masks when out in public
- Other restrictions
- None of the above
- School closures
- Restricted gatherings at (or closed) bars, restaurants, and/or theaters
- Restricted gatherings of a certain number of individuals
- Recommended working from home or not working

Over the past WEEK, on average, how many hours did you sleep per night?

hours per night

Over the past week, how often did you wear a mask (any kind of covering over your mouth and nose) when you're out in public?

- Never
- Sometimes
- Most of the time
- Always
- I did not go out in public this past week

Sections

Monthly COVID-19 Citizen Science Survey

Monthly Surveys

Please answer the following for the period of the past 30 days. Tap next to continue.

What best describes your current main daily activities and/or responsibilities over the past 30 days?

- Working full time
- Working part-time
- Unemployed, laid off, or looking for work
- In school (full- or part-time student)
- Stay-at-home parent or keeping household
- Retired
- Disabled
- Prefer not to state

How much of your working time is currently performed at home?

- 100% of the time
- 75-99% of the time
- 50-74% of the time
- 25-49% of the time
- 1-24% of the time
- None

Has your income changed in the past 30 days?

- Yes, it has increased
- Yes it has declined

No, it is about the same

Prefer not to state

In the past 30 days, by what percentage has your income increased?

%

In the past 30 days, by what percentage has your income declined?

%

In the past 30 days, have you been unemployed?

Yes

No

Prefer not to state

How hard is it for you (and your family) to pay for the very basics like food, rent or mortgage, heating, etc over the past 30 days?

Very hard

Hard

Somewhat hard

Not very hard

Don’t know

Prefer not to state

Did you have difficulty making ends meet over the past 30 days?

Frequently

Occasionally

Hardly ever

Never

Don’t know

Prefer not to state

IN THE PAST WEEK: How many drinks of alcohol (one drink = one standard glass of wine, can of beer, or shot of hard liquor) did you consume?

drinks

Sections

Hospitalization Survey

Monthly Surveys

Have you been hospitalized (had an overnight stay in a hospital) in the past month or since the last time you answered?

Yes

No

How many days did you spend in the hospital over the past 30 days?

days

Have you been to the emergency room or Urgent Care (when you were NOT admitted to the hospital overnight) in the past 30 days or since the last time you answered?

Yes

No

How many times did you go to the emergency room or Urgent Care (when you were NOT admitted to the hospital overnight) in the past 30 days or since the last time you answered?

When were you discharged from the hospital (if more than one time, use most recent)?

MM/DD/YYYY

What was the main reason for your most recent hospitalization (you can look at the papers you received at discharge from the hospital)?

Suspected COVID-19 infection

Asthma

Chronic obstructive pulmonary

Pneumonia

disease

Common flu

Heart attack

Arrhythmias

Other

Please specify the main reason for your hospitalization.

When did you most recently visit the emergency department or Urgent Care?

MM/DD/YYYY

What was the main reason for your most recent emergency department or Urgent Care visit (you can look at the papers you received at discharge from the hospital)?

Suspected COVID-19 infection

Asthma

Chronic obstructive pulmonary disease

Pneumonia

Common flu

Heart attack

Arrhythmias

Other

Please specify the main reason for your most recent emergency department or Urgent Care visit.

Mood Survey

Sections

Monthly Surveys

Over the last 2 weeks, how often have you been bothered by any of the following problems? Tap next to continue.

Little interest or pleasure in doing things.

Not at all
Several days
More than half the days
Nearly every day

Feeling down, depressed, or hopeless.

Not at all
Several days
More than half the days
Nearly every day

Trouble falling or staying asleep, or sleeping too much.

Not at all
Several days
More than half the days
Nearly every day

Feeling tired or having little energy.

Not at all
Several days
More than half the days
Nearly every day

Poor appetite or overeating.

Not at all

Several days

More than half the days

Nearly every day

Feeling bad about yourself - or that you are a failure or have let yourself or your family down.

Not at all

Several days

More than half the days

Nearly every day

Trouble concentrating on things, such as reading the newspaper or watching television.

Not at all

Several days

More than half the days

Nearly every day

Moving or speaking so slowly that other people could have noticed. Or the opposite - being so fidgety or restless that you have been moving around a lot more than usual.

Not at all

Several days

More than half the days

Nearly every day

Anxiety Survey

Sections

Monthly Surveys

Becoming easily annoyed or irritable.

Not at all | Several days
More than half the days | Nearly every day

Over the last two weeks, how often have you been bothered by the following problems? Tap next to continue.

Feeling nervous, anxious, or on edge.

Not at all | Several days
More than half the days | Nearly every day

Not being able to stop or control worrying.

Not at all | Several days
More than half the days | Nearly every day

Worrying too much about different things.

Not at all | Several days
More than half the days | Nearly every day

Trouble relaxing.

Not at all | Several days

More than half the days

Nearly every day

Being so restless that it is hard to sit still.

Not at all

Several days

More than half the days

Nearly every day

Feeling afraid as if something awful might happen.

Not at all

Several days

More than half the days

Nearly every day

Supplementary Appendix 2

Participants of the Covid-19 Citizen Science Study who reported a positive polymerase chain reaction (PCR), antigen, or antibody test prior to enrollment in the study or during their time in the study were called by clinical research coordinators to verify their results and request test documentation to be sent to the study coordinators. In a similar manner to participation in the study, submission of test documentation was entirely voluntary. Thus far, 200 participants who reported prevalent or incident SARS-CoV-2 infections have been called to verify their self-reported results. Of the 93 participants who were reached, 83 verbalized that they would send in their test results, and we have thus far received 52 pieces of documentation to verify self-reported SARS-CoV-2 results. Of the 52 pieces of documentation received, all 52 were either laboratory test results or mandated reporting letters from hospitals/clinics notifying the participant of their PCR or antigen-confirmed SARS-CoV-2 infection.

34 ## Supplementary Tables

Study Week	Proportion of participants who completed at least one survey (%)
1	100%
2	97.3%
3	94.9%
4	92.5%
5	91.8%
6	88.8%
7	88.8%
8	88.8%
9	86.4%
10	84.7%
11	83.1%
12	82.1%
13	83.3%
14	83.1%
15	82.8%
16	82.3%
17	88.7%
18	86.7%
19	86.8%
20	86.2%
21	85.9%
22	88.1%
23	88.8%
24	88.4%
25	89.2%
26	90.2%
27	93.0%
28	100%

35 **Supplementary Table 1.** Mean proportion of participants who completed at least one health
36 survey each week.
37

Study Month	Proportion of participants who completed at least one survey (%)
1	100%
2	98.2%
3	95.6%
4	96.9%
5	97.4%
6	98.5%
7	100%

Supplementary Table 2. Mean proportion of participants who completed at least one health survey each month.

	Odds Ratio	95% CI	p-value	Group p-value
Age (years)	0.98	0.97, 1.00	0.014	
Female Biological Sex	0.95	0.59, 1.54	0.84	
Race/Ethnicity				
White	reference			0.40*
Black	2.96	0.71, 12.29	0.13	0.52†
Hispanic (any race)	1.19	0.53, 2.65	0.67	
Other (including multiracial)	1.69	0.53, 5.40	0.38	
MacArthur Subjective Social Status Ladder	0.92	0.82, 1.04	0.19	
Alcoholic drinks per week, last 4-21 days	0.97	0.93, 1.00	0.07	
Number of contacts (per 10), last 4-21 days	1.11	1.02, 1.21	0.012	
Number of events with 10 or more people (per 10), last 4-21 days	1.26	1.07, 1.48	0.006	
Number of visits to movie theaters (per 10), last 4-21 days	2.00	0.97, 4.11	0.06	
Number of visits to restaurants (per 10), last 4-21 days	1.85	1.37, 2.49	<0.001	
Weeks since study start (linear)	1.04	1.01, 1.07	0.017	

Supplementary Table 3. Backward stepwise logistic model for incident SARS-CoV-2 infection using retention criterion of $p<0.1$ with standard errors clustered on participants.

* overall heterogeneity

† heterogeneity of non-reference levels

linear trend

47

	Odds Ratio	95% CI	p-value	Group p-value
Age (years)	0.98	0.96, 0.99	0.008	
Female Biological Sex	0.81	0.49, 1.34	0.42	
Race/Ethnicity				
White	reference			0.43*
Black	3.00	0.72, 12.53	0.13	0.56†
Hispanic (any race)	1.35	0.64, 2.86	0.43	
Other (including multiracial)	1.19	0.28, 4.97	0.81	
MacArthur Subjective Social Status Ladder	0.93	0.82, 1.05	0.24	
Alcoholic drinks per week, last 4-21 days	0.97	0.94, 1.00	0.06	
Number of contacts (per 10), last 4-21 days	1.10	1.00, 1.21	0.04	
Number of events with 10 or more people (per 10), last 4-21 days	1.29	1.09, 1.53	0.003	
Number of visits to movie theaters (per 10), last 4-21 days	1.99	0.97, 4.08	0.059	
Number of visits to restaurants (per 10), last 4-21 days	2.31	1.46, 3.63	<0.001	
Weeks since study start (linear)	1.04	1.01, 1.07	0.008	

48 **Supplementary Table 4.** Backward stepwise logistic model for incident SARS-CoV-2 infection
49 using retention criterion of p<0.1 with standard errors clustered on FIPS county-level codes
50 (using US participants only).

51 * overall heterogeneity

52 † heterogeneity of non-reference levels

53 # linear trend

54

	Odds Ratio	**95% CI**	**p-value**	**Group p-value**
Age (years)	0.98	0.97, 0.99	0.007	
Female Biological Sex	0.88	0.55, 1.42	0.60	
Race/Ethnicity				
White	reference			0.39*
Black	2.91	0.70, 12.09	0.14	0.55†
Hispanic (any race)	1.23	0.55, 2.75	0.62	
Other (including multiracial)	1.74	0.54, 5.57	0.35	
MacArthur Subjective Social Status Ladder	0.92	0.81, 1.04	0.17	
Alcoholic drinks per week, last 4-21 days	0.97	0.93, 1.00	0.07	
Number of contacts (per 10), last 4-21 days	1.12	1.03, 1.21	0.008	
Number of events with 10 or more people (per 10), last 4-21 days	1.29	1.09, 1.52	0.003	
Number of visits to movie theaters (per 10), last 4-21 days	1.98	0.95, 4.09	0.07	
Number of visits to restaurants (per 10), last 4-21 days	1.83	1.36, 2.47	<0.001	
Weeks since study start (linear)	1.04	1.01, 1.07	0.015	

Supplementary Table 5. Backward stepwise logistic model for incident SARS-CoV-2 infection using retention criterion of p<0.1 with standard errors clustered on zip codes (using US participants only).

* overall heterogeneity

† heterogeneity of non-reference levels

\# linear trend

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract (*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	1
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	5-6
Objectives	3	State specific objectives, including any prespecified hypotheses	6
Methods			
Study design	4	Present key elements of study design early in the paper	6-8
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	6-8
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up (*b*) For matched studies, give matching criteria and number of exposed and unexposed	7
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	7-8
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	7-8
Bias	9	Describe any efforts to address potential sources of bias	6-7
Study size	10	Explain how the study size was arrived at	6-7
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	8-9
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding (*b*) Describe any methods used to examine subgroups and interactions (*c*) Explain how missing data were addressed (*d*) If applicable, explain how loss to follow-up was addressed (*e*) Describe any sensitivity analyses	8-9
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed (b) Give reasons for non-participation at each stage (c) Consider use of a flow diagram	9-10
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders (b) Indicate number of participants with missing data for each variable of interest (c) Summarise follow-up time (eg, average and total amount)	9-10
Outcome data	15*	Report numbers of outcome events or summary measures over time	9-10

Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included (*b*) Report category boundaries when continuous variables were categorized (*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	9-10
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	10
Discussion			
Key results	18	Summarise key results with reference to study objectives	10-11
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	13-14
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	14
Generalisability	21	Discuss the generalisability (external validity) of the study results	13-14
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	20

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.